Fall 2013

Efficient Spectral-Element Methods For Acoustic Scattering And Related Problems

Ying He
Purdue University

Graduate School ETD Form 9
(Revised 12/07)

PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis/Dissertation Acceptance

This is to certify that the thesis/dissertation prepared

By YING HE

Entitled Efficient Spectral-Element Methods for Acoustic Scattering and Related Problems

For the degree of Doctor of Philosophy

Is approved by the final examining committee:

Jie Shen
Chair

Patricia Bauman

Peijun Li

Robert Skeel

To the best of my knowledge and as understood by the student in the *Research Integrity and Copyright Disclaimer (Graduate School Form 20)*, this thesis/dissertation adheres to the provisions of Purdue University's "Policy on Integrity in Research" and the use of copyrighted material.

Approved by Major Professor(s): Jie Shen

Approved by: David Goldberg 10/04/2013
Head of the Graduate Program Date

EFFICIENT SPECTRAL-ELEMENT METHODS FOR ACOUSTIC SCATTERING AND RELATED PROBLEMS

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Ying He

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

December 2013

Purdue University

West Lafayette, Indiana

To my husband and parents, for their endless love and support.

ACKNOWLEDGMENTS

I owe my gratitude to all those people who have made this dissertation possible and because of whom my graduation experience has been one that I will cherish forever.

Foremost, I would like to express my sincere gratitude to my advisor, Prof. Jie Shen, for leading me to today's achievements. His perpetual enthusiasm in research and meticulous scholarship have influenced me immensely in the course of performing research. His extensive knowledge and insightful vision have been the source of inspiration for me throughout my research work. I really appreciate his belief in my potential, which allows me to explore problems extensively.

I would like to thank my other dissertation committee members, Prof. Robert D. Skeel, Prof. Patricia E. Bauman and Prof. Peijun Li, for taking their time to help me improve this dissertation.

I would like to express my gratitude to all the members of my research group, schoolmates, and professors at Purdue University, for helping me in many ways during these years. I would also like to extend a general thank you to all teachers I have had throughout my life. I am where I am today because of them.

Finally, I would like to thank my family, particularly my father Bin He, my mother Zhiqiong Ren, and my husband Changhui Lin, for their unconditional support throughout my life and career.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

He, Ying Ph.D., Purdue University, December 2013. Efficient Spectral-Element Methods for Acoustic Scattering and Related Problems. Major Professor: Jie Shen .

This dissertation focuses on the development of high-order numerical methods for acoustic and electromagnetic scattering problems, and nonlinear fluid-structure interaction problems.

For the scattering problems, two cases are considered: 1) the scattering from a doubly layered periodic structure; and 2) the scattering from doubly layered, unbounded rough surface. For both cases, we first apply the transformed field expansion (TFE) method to reduce the two-dimensional Helmholtz equation with complex scattering surface into a successive sequence of the transmission problems with a plane interface. Then, we use Fourier-Spectral method in the periodic structure problem and Hermite-Spectral method in the unbounded rough surface problem to reduce the two-dimensional problems into a sequence of one-dimensional problems, which can then be efficiently solved by a Legendre-Galerkin method.

In order for TFE method to work well, the scattering surface has to be a sufficiently small and smooth deformation of a plane surface. To deal with scattering problems from a non-smooth surface, we also develop a high-order spectral-element method which is more robust than the TFE method, but is computationally more expensive.

We also consider the non-linear fluid-structure interaction problem, and develop a class of monolithic pressure-correction schemes, based on the standard pressure-correction and rotational pressure-correction schemes. The main advantage of these schemes is that they only require solving a pressure Poisson equation and a linear coupled elliptic equation at each time step. Hence, they are computationally very efficient. Furthermore, we prove that the proposed schemes are unconditionally stable.

CHAPTER 1. OVERVIEW OF THE DISSERTATION

Most acoustic, electromagnetic and fluid dynamics problems in the real world, such as scattering, radiation, waveguide, fluid-structure interaction, etc., are not analytically solvable. However, solutions of many of these problems are governed by partial differential equations (PDEs), and well designed computational techniques can overcome the inability to derive closed form solutions. This dissertation focuses on using *high-order numerical methods* to solve a class of acoustic, electromagnetic and related fluid dynamics problems.

The interaction of acoustic and electromagnetic waves with layered medium plays an important role in a wide range of problems of scientific and technological interest. We are mainly concerned with the scattering problems from the layered medium with *periodic structure* and *unbounded rough surface*. Another important topic of this dissertation is to design efficient numerical schemes for fluid-structure interaction (FSI) problems. The main contents of thesis are briefly described below.

1.1 Scattering by Periodic Structure

The scattering of electromagnetic or acoustic waves by a periodic structure has found applications in many engineering problems, from remote sensing [1], to grating couplers [2–4], to nanostructures [5]. We focus on developing high-order numerical methods for the interaction of time-harmonic electromagnetic waves incident upon a periodic doubly layered dielectric material with a sharply defined, irregular interface. While we only consider the simplified model of a two-dimensional structure, the core of the algorithms will remain the same for three-dimensional problems governed by the full Maxwell's equations.

We propose two numerical schemes for the case of periodic structure: *Transformed Field Expansion* (TFE) approach and high-order *Spectral Element Method* (SEM) in combination with a DtN technique.

Transformed Field Expansion (TFE). TFE approach is a Boundary Perturbation Method (BPM) for the numerical simulation of scattering returns from an irregularly shaped, periodic, doubly layered medium. For such problems, one compelling choice is a surface integral method [6] (e.g., Boundary Integral Methods–BIM–or Boundary Element Methods–BEM) which only requires a discretization of the layer *interface* (rather than the whole structure) and which, due to the choice of the Green's function, enforces the far-field boundary condition exactly. While these methods can deliver high-accuracy simulations with greatly reduced operation counts, there are several difficulties which need to be addressed. An alternative to BIM/BEM is the Boundary Perturbation Method, and two popular approaches are the "Method of Field Expansions" (FE) due to Bruno & Reitich [7–9] and the "Method of Operator Expansions" (OE) of Milder [10–15]. These methods are very appealing as they only solve for *surface* unknowns, thereby enjoying favorable operation counts of surface integral methods, while avoiding the subtle quadrature rules, dense linear systems, and required matrix-vector product accelerations described above. However, Nicholls & Reitich showed that these algorithms depend upon strong cancellations (e.g., differences of extremely large quantities to produce order one results) for their convergence, and they often lead to ill-conditioned numerics. We refer the interested reader to [16–18] for a full description of these phenomena.

To overcome the ill-conditioning, Nicholls & Reitich described an alternative Boundary Perturbation algorithm, the "Method of Transformed Field Expansions" (TFE), which does *not* rely on strong cancellations for its convergence, and where a change of variables was done first to flatten the shape of the scattering surface and then followed by the boundary perturbation technique.

In Chapter 2, we construct a non-trivial extension of the TFE algorithm to the case of doubly layered periodic structure. In particular, we develop an efficient and stable spectral method for the two-dimensional Helmholtz equation in a two-layered periodic domain, which has to be solved repeatedly in the TFE algorithm.

Spectral Element Method (SEM). In Chapter 3, we present a high-order spectral element method in combination with a transparent boundary condition for the numerical simulation of scattering returns from an irregularly shaped, periodic, doubly layered medium. Time-harmonic scattering problem is considered and represented by the Helmholtz equation, whose solution is approximated by a tensor product basis of the one-dimensional Lagrange interpolation polynomials based on the Legendre-Gauss-Lobatto points. For SEM methods, we refer the interested reader to [16–18] for a detailed description. The advantage of applying the SEM rather than the TFE approach is that, the geometric structures can be accurately represented by body-fitted quadrilateral elements, so the SEM approach is more suitable for non-smooth interface structure. To solve the resulting large sparse system, a GMRES iteration technique is used. Like many other existing iteration methods for solving Helmholtz equations, a good preconditioner is crucial for accelerating the convergence of the GMRES iterations. However, the indefiniteness of Helmholtz problem makes it difficult to design an efficient preconditioner, in particular, for moderate to large wave numbers.

We demonstrate computational results for the geometries with smooth flat structures, smooth curved structures, and nonsmooth interface structures. We validate our results provided with convergence studies in comparison with exact solutions, TFE solutions, and the energy defect measure. Our results show exponential convergence as increasing the approximation order.

Of particular importance to this approach, the SEM method was generalized to precisely the case we treat here (a doubly layered, two-dimensional material) by two of the authors [19]. We use this algorithm as a test for the numerical approach

we advocate in this dissertation, and with it we can illuminate the strengths and shortcomings of this new SEM approach.

1.2 Scattering by Unbounded Rough Surface

The phenomenon of acoustic and electromagnetic scattering by unbounded rough surfaces has received much attention from both the engineering and mathematical communities for its important applications in a wide range of scientific areas. An unbounded rough surface is referred to as a non-local perturbation of an infinite plane surface such that the whole surface lies within a finite distance of the original plane. Due to the non-locally perturbed scattering surfaces, precise modeling and accurate computing present challenging mathematical and computational questions.

A considerable amount of efforts have been denoted to the rough surface scattering problems by using approximate, asymptotic, or statistical methods, e.g., the reviews and monographs by Ogilvy [20], Voronovich [21], Saillard and Sentenac [22], Warnick and Chew [23], DeSanto [24], Elfouhaily and Guerin [25], and references cited therein. Despite the large amount of work done so far, we are not aware of any efficient and accurate numerical method for solving the scattering problem by unbounded rough surfaces.

In Chapter 4, we are concerned with the numerical solution for the unbounded rough surface scattering problem. Specifically, we study the acoustic wave propagation problem of the two-dimensional Helmholtz equation with an unbounded penetrable scattering surface.

Under the assumption that scattering rough surface is a sufficiently small and smooth deformation of a plane surface, we use the transformed field expansion to reduce the two-dimensional Helmholtz equation with complex scattering surface into a successive sequence of the transmission problems with a plane interface and piecewise constant wave numbers. We then use Hermite functions, which form an orthonormal basis of $L^2(\mathbb{R})$ and are eigenfunctions of continuous Fourier transform, to handle the

difficulty from the infinite domain in horizontal direction, and further reduce the two-dimensional transmission problems into fully decoupled one-dimensional two-point boundary value problems, which can be efficiently solved by a Legendre-Galerkin method. We present numerical examples for both the rough surface scattering and the plane surface scattering, where the analytic solution is available.

1.3 Non-linear Fluid-structure Interaction Problem

Fluid-Structure Interaction (FSI) plays an important role in many scientific and engineering applications, e.g., design of some engineering systems (e.g., aircraft and bridges), blood flow in human arteries, etc. It has been extensively studied in recent years both analytically and computationally (cf., [26–28] and the references therein).

From the physical point of view, the FSI problem has little connection with the scattering problem, which is the main focus of this dissertation. However, our numerical schemes lead to, at each time step, linear systems that have essentially the same structure as those from the scattering problems and can be solved by using similar procedures.

In Chapter 5, we consider a simple model of the FSI problem where the movement of the interface is assumed infinitesimal so the interface is treated as fixed. This nonlinear FSI problem captures many of the essential difficulties of the more general FSI problems with moving interface, and its well-posedness has been well studied in [29].

There are two main approaches, *monolithic* (cf., for instance, [30–32]) and *partitioned* (cf., for instance, [33–35]) methods, for solving FSI problems numerically. Compared with partitioned method, the monolithic methods usually have good stability properties, but at each time step, a nonlinear coupled system has to be solved, and due to the presence of the pressure in the coupled system, it is usually difficult to design effective iterative scheme to solve the nonlinear coupled system.

For fluid problems, an effective approach to decouple the computation of the pressure from that of the velocity is to use a so called projection type method, originally proposed by Chorin and Temam in the late 60's. A comprehensive review on various projection type methods can be found in [36]. Naturally, many authors have considered to employ a projection type method for the FSI problem (cf., for instance, [37]). However, a main difficulty in the design of a projection method is what boundary condition to use for the pressure at the interface. It is well known that a proper boundary condition, at the Dirichlet part of the boundary, for the pressure Poisson equation in a projection type method is the homogeneous Neumann boundary condition. Indeed, most existing projection type schemes for FSI problem also use, explicitly or implicitly, Neumann type boundary condition for the pressure Poisson equation at the interface. However, we are not aware of any rigorous proof of unconditional stability for any projection type scheme applied to the FSI problem.

We construct several monolithic schemes based on a pressure-correction approach for the FSI model with fixed interface. Our schemes will be computationally very efficient. More precisely, in the first step of our schemes, we solve a coupled, but elliptic, system for an intermediate fluid velocity and the structure displacement, then in the second step, we solve a Poisson equation for the fluid pressure and obtain the fluid velocity with a simple correction. We shall also prove rigorously that these schemes are unconditionally stable.

1.4 Extensions and Future Work

In Chapter 6, two possible extensions are discussed. One is the scattering problem from open cavity; the other is the nonlinear fluid-structure interaction with moving interface. We present some basic ideas about the numerical approximation for those two challenging problems.

CHAPTER 2. AN EFFICIENT AND STABLE SPECTRAL METHOD FOR ELECTROMAGNETIC SCATTERING FROM A LAYERED PERIODIC STRUCTURE

In this chapter, we focus on developing a stable and high-order numerical method for time-harmonic electromagnetic waves incident upon a doubly periodic layered dielectric media with irregular interface. We describe a Boundary Perturbation Method for this problem which avoids not only the need for specialized quadrature rules but also the dense linear systems which are characteristic of Boundary Integral/Element Methods. Moreover, it is a provably stable algorithm as opposed to other Boundary Perturbation approaches such as Bruno & Reitich's "Method of Field Expansions" or Milder's "Method of Operator Expansions". Our spectrally accurate approach is a natural extension of the "Method of Transformed Field Expansions", originally described by Nicholls & Reitich (and later refined to other geometries by the authors) in the single-layer case.

2.1 Introduction

The interaction of acoustic and electromagnetic waves with periodic structures plays an important role in a wide range of problems of scientific and technological interest. From grating couplers [2–4] to nanostructures [5] to remote sensing [38], the ability to simulate in a robust and accurate way the fields generated by such structures is of crucial importance to researchers from many disciplines. In this contribution we focus upon the stable and high–order numerical simulation of the interaction of time–harmonic electromagnetic waves incident upon a periodic doubly layered dielectric material with sharp, irregular interface. While we focus on the simplified model of a

two–dimensional structure, the core of the algorithm will remain the same for a fully three–dimensional simulation governed by the full Maxwell's equations.

In this chapter we describe a Boundary Perturbation Method (BPM) for the numerical simulation of scattering returns from an irregularly shaped, periodic, doubly layered medium. We focus upon periodic structures as they arise from a large number of engineering applications, however, this choice does simplify our numerical approach (e.g., we may use the Discrete Fourier Transform to approximate Fourier coefficients). However, we note that this simplification is also realized for competing methods as well. For such problems *surface* methods are preferred as a discretization of the interface alone significantly reduces the number of unknowns to be recovered. However, such methods face a number of drawbacks.

One compelling choice is a surface integral method [6] (e.g., Boundary Integral Methods–BIM–or Boundary Element Methods–BEM) which only require a discretization of the layer *interface* (rather than the whole structure) and which, due to the choice of the Green's function, enforce the far–field boundary condition exactly. While these methods can deliver high–accuracy simulations with greatly reduced operation counts, there are several difficulties which need to be addressed. First, high–order simulations can only be realized with specially designed quadrature rules which respect the singularities in the Green's function (and its derivative, in certain formulations). Additionally, BIM/BEM typically give rise to dense linear systems to be solved which require carefully designed preconditioned iterative methods (with accelerated matrix–vector products, e.g., by the Fast–Multipole Method [39]) for configurations of engineering interest.

An alternative to a BIM/BEM is a Boundary Perturbation Method and two popular approaches are the "Method of Field Expansions" (FE) due to Bruno & Reitich [7–9] and the "Method of Operator Expansions" (OE) of Milder [10–15]. These methods are very appealing as they posit *surface* unknowns thereby enjoying the favorable operation counts of surface integral methods, while avoiding the subtle quadrature rules, dense linear systems, and required matrix–vector product acceler-

ations described above. However, Nicholls & Reitich showed that these algorithms depend upon strong cancellations (e.g., differences of extremely large quantities to produce order one results) for their convergence which results in ill–conditioned numerics. We refer the interested reader to [16–18] for a full description of these phenomena.

In addition to these results, Nicholls & Reitich described an alternative Boundary Perturbation algorithm, the "Method of Transformed Field Expansions" (TFE), which does *not* rely on strong cancellations for its convergence. In fact, the resulting recursions can be used for a *direct*, *rigorous* demonstration of the strong convergence of the relevant perturbation expansions in an appropriate function space. Furthermore, these formulas were implemented to reveal a stable and highly accurate numerical scheme for the simulation of scattering returns by periodic gratings. This work was generalized by the authors to the case of irregular bounded obstacles in two [40] and three dimensions [41], and even resulted in a rigorous numerical analysis of the method [42]. In this contribution, we construct a highly non-trivial extention to the case of periodic gratings separating two materials of different dielectric constants. Here, of course, one must be concerned not only with a reflected field and its far–field boundary condition, but also with a transmitted field which satisfies a different condition at infinity.

The organization of this chapter is as follows: In § 2.2 we recall the governing equations of an electromagnetic field incident upon a periodic, two–dimensional irregular grating. In § 2.3 we define a change of variables which significantly enhances the conditioning properties of our numerical scheme resulting in the "Method of Transformed Field Expansions." We discuss a Legendre Galerkin method to solve the resulting two–point boundary value problem in § 2.4 and present extensive numerical results in § 2.5.

2.2 Governing Equations

We consider the problem of simulating the scattering of electromagnetic waves in a layered periodic structure. More precisely, we consider two domains

$$\Omega^+ := \{y > g(x)\}, \quad \Omega^- := \{y < g(x)\},$$

where $y = g(x)$ is the shape of the d–periodic interface (see Figure 3.1). These

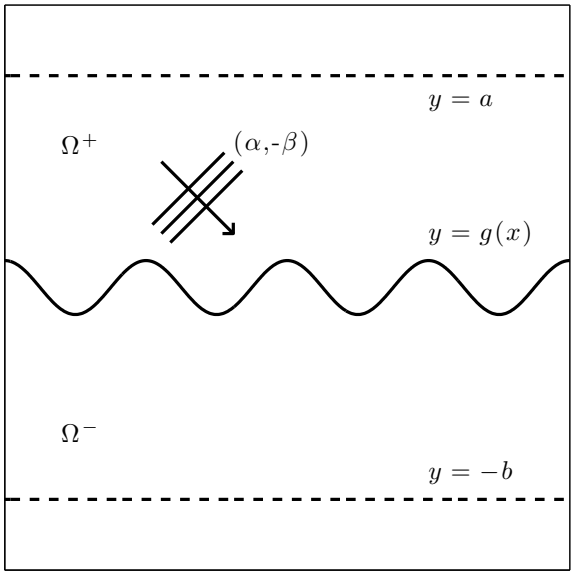


Figure 2.1. Geometric illustration of the problem

regions are filled with materials of dielectric constants ϵ^+ and ϵ^-, respectively. The permeability in each domain is assumed to be μ_0, that of the vacuum.

The grating is illuminated by time–harmonic plane–wave radiation

$$\tilde{\mathbf{E}}^{\mathbf{i}} = \mathbf{A}e^{i\alpha x - i\beta y}e^{-i\omega t}, \quad \tilde{\mathbf{H}}^{\mathbf{i}} = \mathbf{B}e^{i\alpha x - i\beta y}e^{-i\omega t},$$

which will be scattered both above and below the surface. This gives rise to reduced total fields

$$\mathbf{E} = \mathbf{E}^{\mathbf{i}} + \mathbf{E}^+, \quad \mathbf{H} = \mathbf{H}^{\mathbf{i}} + \mathbf{H}^+, \quad y > g(x),$$
$$\mathbf{E} = \mathbf{E}^-, \quad \mathbf{H} = \mathbf{H}^-, \quad y < g(x),$$

where, e.g.,

$$\mathbf{E} = \mathbf{E}(x,y) := \tilde{\mathbf{E}}(x,y,t)e^{i\omega t}, \quad \mathbf{H} = \mathbf{H}(x,y) := \tilde{\mathbf{H}}(x,y,t)e^{i\omega t},$$

if $\{\tilde{\mathbf{E}}, \tilde{\mathbf{H}}\}$ are the unreduced, time dependent fields. The incident, reflected, refracted, and total electric and magnetic fields all satisfy the time–harmonic Maxwell's equations:

$$\nabla \times \mathbf{E} = i\omega\mu_0\mathbf{H}, \quad \operatorname{div}[\mathbf{E}] = 0, \tag{2.2.1a}$$

$$\nabla \times \mathbf{H} = i\omega\epsilon\mathbf{E}, \quad \operatorname{div}[\mathbf{H}] = 0, \tag{2.2.1b}$$

where $\epsilon = \epsilon^{\pm}$ depending upon the domain of definition [6]. At the grating surface the total fields satisfy the transmission conditions

$$N \times \left(\mathbf{E^i} + \mathbf{E}^+ - \mathbf{E}^-\right) = 0, \quad N \times \left(\mathbf{H^i} + \mathbf{H}^+ - \mathbf{H}^-\right) = 0, \tag{2.2.2}$$

where $N = (-\partial_x g(x), 1)^T$ is a normal vector. Finally, the periodicity of the grating enforces the *quasi-periodicity* of the fields

$$\mathbf{E}(x+d,y) = e^{i\alpha d}\mathbf{E}(x,y), \quad \mathbf{H}(x+d,y) = e^{i\alpha d}\mathbf{H}(x,y),$$

and the scattered waves must be *outgoing*.

It is not difficult to show that if both the grating shape and incident radiation are independent of z then so are $\mathbf{E}$ and $\mathbf{H}$ [43]. In this case the time–harmonic Maxwell's equations (2.2.1) reduce to the Helmholtz equation

$$\Delta u + (k^{\pm})^2 u = 0,$$

where $k^{\pm} := \omega\sqrt{\mu_0\epsilon^{\pm}}$, and $u = u(x,y)$ is either $E^{\pm,3}$ (transverse electric—TE—component) or $H^{\pm,3}$ (transverse magnetic—TM—component). Furthermore, the z–components of the conditions in (2.2.2) read

$$0 = E^3 = E^{+,3} + E^{i,3} - E^{-,3}$$

and

$$0 = \partial_N E^3 = [\partial_y - (\partial_x g)\partial_x]\,(E^3) = [\partial_y - (\partial_x g)\partial_x]\,(E^{+,3} + E^{i,3} - E^{-,3}).$$

Writing in coordinates and simplifying we find

$$u^+(x,g(x)) - u^-(x,g(x)) = -e^{i\alpha x - i\beta g(x)},$$
$$\partial_N u^+(x,g(x)) - \sigma^2 \partial_N u^-(x,g(x)) = \left((i\beta) + (i\alpha)\partial_x g(x)\right) e^{i\alpha x - i\beta g(x)},$$

where $\sigma^2 = 1$ for the TE mode, while

$$\sigma^2 = \frac{\epsilon^+}{\epsilon^-} = \left(\frac{k^+}{k^-}\right)^2,$$

for the TM mode. Thus, for the TE mode, the governing equations we consider are

$$\Delta u^+ + (k^+)^2 u^+ = 0, \qquad y > g(x), \tag{2.2.3a}$$
$$\mathrm{OWC}[u^+] = 0, \qquad y \to \infty, \tag{2.2.3b}$$
$$\Delta u^- + (k^-)^2 u^- = 0, \qquad y < g(x) \tag{2.2.3c}$$
$$\mathrm{OWC}[u^-] = 0, \qquad y \to -\infty, \tag{2.2.3d}$$
$$u^+ - u^- = -\phi(x), \qquad y = g(x), \tag{2.2.3e}$$
$$\partial_N u^+ - \partial_N u^- = \left((i\beta) + (i\alpha)\partial_x g(x)\right)\phi(x), \qquad y = g(x), \tag{2.2.3f}$$
$$u^\pm(x+d,y) = e^{i\alpha d} u^\pm(x,y), \tag{2.2.3g}$$

where

$$\phi(x) := e^{i\alpha x - i\beta g(x)}, \tag{2.2.3h}$$

and we make the "outgoing wave condition" (OWC) operators more precise presently.

For the far–field boundary conditions consider the hyperplanes $\{y = a\}$, $\{y = -b\}$ where $a, b > |g|_{L^\infty}$. The augmented system of governing equations

$$\Delta u^+ + (k^+)^2 u^+ = 0, \qquad g(x) < y < a, \tag{2.2.4a}$$

$$\partial_y u^+ = \partial_y v^+, \qquad y = a, \tag{2.2.4b}$$

$$u^+ = v^+, \qquad y = a, \tag{2.2.4c}$$

$$\Delta v^+ + (k^+)^2 v^+ = 0, \qquad y > a, \tag{2.2.4d}$$

$$\mathrm{OWC}[v^+] = 0, \qquad y \to \infty, \tag{2.2.4e}$$

$$\Delta u^- + (k^-)^2 u^- = 0, \qquad -b < y < g(x), \tag{2.2.4f}$$

$$\partial_y u^- = \partial_y v^-, \qquad y = -b, \tag{2.2.4g}$$

$$u^- = v^-, \qquad y = -b, \tag{2.2.4h}$$

$$\Delta v^- + (k^-)^2 v^- = 0, \qquad y < -b, \tag{2.2.4i}$$

$$\mathrm{OWC}[v^-] = 0, \qquad y \to -\infty, \tag{2.2.4j}$$

$$u^+ - u^- = -\phi(x), \qquad y = g(x), \tag{2.2.4k}$$

$$\partial_N u^+ - \partial_N u^- = \left((i\beta) + (i\alpha)\partial_x g(x)\right)\phi(x), \qquad y = g(x), \tag{2.2.4l}$$

$$u^\pm(x + d, y) = e^{i\alpha d} u^\pm(x, y), \tag{2.2.4m}$$

$$v^\pm(x + d, y) = e^{i\alpha d} v^\pm(x, y), \tag{2.2.4n}$$

are equivalent to (3.2.7). To make the far–field boundary condition more precise we note that solutions of (3.2.13)–(3.2.14) are

$$v^+(x, y) = \sum_{p=-\infty}^{\infty} \hat{\psi}_p e^{i\alpha_p x + i\beta_p^+ (y-a)},$$

where

$$\alpha_p := \alpha + (2\pi/d)p, \quad \beta_p^\pm := \begin{cases} \sqrt{(k^\pm)^2 - \alpha_p^2} & p \in U^\pm \\ i\sqrt{\alpha_p^2 - (k^\pm)^2} & p \notin U^\pm \end{cases},$$

$$U^\pm := \left\{ p \in \mathbf{Z} \mid (k^\pm)^2 - \alpha_p^2 > 0 \right\},$$

$\mathbf{Z}$ are the integers, and

$$\psi(x) := u^+(x, a) = \sum_{p=-\infty}^{\infty} \hat{\psi}_p e^{i\alpha_p x}.$$

Similarly, solutions of (2.2.4i)–(2.2.4j) are

$$v^-(x,y) = \sum_{p=-\infty}^{\infty} \hat{\zeta}_p e^{i\alpha_p x - i\beta_p^-(y+b)},$$

where

$$\zeta(x) := u^-(x,-b) = \sum_{p=-\infty}^{\infty} \hat{\zeta}_p e^{i\alpha_p x}.$$

To close the set of equations for u^+ we simply need to produce $\partial_y v^+$ in (3.2.11):

$$\partial_y v^+(x,a) = \sum_{p=-\infty}^{\infty} (i\beta_p^+)\hat{\psi}_p e^{i\alpha_p x} =: T^+[\psi] = T^+[u^+(x,a)].$$

A similar analysis at $y = -b$ yields an operator

$$T^-[\zeta] := \sum_{p=-\infty}^{\infty} (-i\beta_p^-)\hat{\zeta}_p e^{i\alpha_p x},$$

and the system (2.2.4) can be *equivalently* restated as

$$\Delta u^+ + (k^+)^2 u^+ = 0, \qquad g(x) < y < a, \tag{2.2.5a}$$

$$\partial_y u^+ - T^+[u^+] = 0, \qquad y = a, \tag{2.2.5b}$$

$$\Delta u^- + (k^-)^2 u^- = 0, \qquad -b < y < g(x), \tag{2.2.5c}$$

$$\partial_y u^- - T^-[u^-] = 0, \qquad y = -b, \tag{2.2.5d}$$

$$u^+ - u^- = -\phi(x), \qquad y = g(x), \tag{2.2.5e}$$

$$\partial_N u^+ - \partial_N u^- = \left((i\beta) + (i\alpha)\partial_x g(x)\right)\phi(x), \qquad y = g(x), \tag{2.2.5f}$$

$$u^\pm(x+d,y) = e^{i\alpha d} u^\pm(x,y). \tag{2.2.5g}$$

2.3 Transformed Field Expansion

As has been demonstrated in previous publications on Boundary Perturbation algorithms for electromagnetic scattering [16–18], the transformed field expansion (TFE) method can dramatically improve the conditioning of the resulting recursions. The TFE method consists of two essential steps: (i) "domain flattening" through a simple change of variables; and (ii) boundary perturbation. We now describe the two steps in detail.

2.3.1 Change of Variables

We define

$$\begin{aligned} x' &= x, \\ y' &= a\left(\frac{y-g}{a-g}\right), \quad g < y < a, \\ y'' &= b\left(\frac{y-g}{b+g}\right), \quad -b < y < g, \end{aligned}$$

which gives rise to the differentiation rules:

$$\begin{aligned} (a-g)\partial_x &= (a-g)\partial_{x'} - (\partial_{x'}g)(a-y')\partial_{y'}, \\ (a-g)\partial_y &= a\partial_{y'}, \end{aligned}$$

for $g(x) < y < a$, and

$$\begin{aligned} (b+g)\partial_x &= (b+g)\partial_{x'} - (\partial_{x'}g)(b+y'')\partial_{y''}, \\ (b+g)\partial_y &= b\partial_{y''}, \end{aligned}$$

for $-b < y < g(x)$. With this change of variables (3.2.25) becomes

$$\left(\partial_{x'}^2 + \partial_{y'}^2\right) u^+(x',y') + (k^+)^2 u^+(x',y') = F^+(x',y'), \qquad 0 < y' < a, \tag{2.3.1a}$$

$$\partial_{y'} u^+(x',a) - T^+[u^+(x',a)] = J^+(x'), \tag{2.3.1b}$$

$$\left(\partial_{x'}^2 + \partial_{y''}^2\right) u^-(x',y'') + (k^-)^2 u^-(x',y'') = F^-(x',y''), \qquad -b < y'' < 0, \tag{2.3.1c}$$

$$\partial_{y''} u^-(x',-b) - T^-[u^-(x',-b)] = J^-(x'), \tag{2.3.1d}$$

$$u^+(x',0) - u^-(x',0) = -\phi(x'), \tag{2.3.1e}$$

$$\partial_{y'} u^+(x',0) - \partial_{y''} u^-(x',0) = Q(x'). \tag{2.3.1f}$$

In these equations

$$F^\pm(x',y') = \partial_{x'} F_x^\pm(x',y') + \partial_{y'} F_y^\pm(x',y') + F_h^\pm(x',y'), \tag{2.3.1g}$$

where

$$F_x^+ = \frac{2}{a} g \partial_{x'} u^+ - \frac{1}{a^2} g^2 \partial_{x'} u^+ + \frac{a-y'}{a}(\partial_{x'} g)\partial_{y'} u^+ - \frac{a-y'}{a^2} g (\partial_{x'} g)\partial_{y'} u^+, \tag{2.3.1h}$$

$$F_y^+ = \frac{a-y'}{a}(\partial_{x'}g)\partial_{x'}u^+ - \frac{a-y'}{a^2}g(\partial_{x'}g)\partial_{x'}u^+ - \frac{(a-y')^2}{a^2}(\partial_{x'}g)^2\partial_{y'}u^+, \qquad (2.3.1\text{i})$$

and

$$F_h^+ = \frac{1}{a^2}g(\partial_{x'}g)\partial_{x'}u^+ - \frac{1}{a}(\partial_{x'}g)\partial_{x'}u^+ + \frac{a-y'}{a^2}(\partial_{x'}g)^2\partial_{y'}u^+ + (k^+)^2\frac{2}{a}gu^+ - (k^+)^2\frac{1}{a^2}g^2u^+, \qquad (2.3.1\text{j})$$

and

$$F_x^- = -\frac{2}{b}g\partial_{x'}u^- - \frac{1}{b^2}g^2\partial_{x'}u^- + \frac{b+y'}{b}(\partial_{x'}g)\partial_{y'}u^- + \frac{b+y'}{b^2}g(\partial_{x'}g)\partial_{y'}u^-, \qquad (2.3.1\text{k})$$

$$F_y^- = \frac{b+y'}{b}(\partial_{x'}g)\partial_{x'}u^- + \frac{b+y'}{b^2}g(\partial_{x'}g)\partial_{x'}u^- - \frac{(b+y')^2}{b^2}(\partial_{x'}g)^2\partial_{y'}u^-, \qquad (2.3.1\text{l})$$

and

$$F_h^- = \frac{1}{b}(\partial_{x'}g)\partial_{x'}u^- + \frac{1}{b^2}g(\partial_{x'}g)\partial_{x'}u^- - \frac{b+y'}{b^2}(\partial_{x'}g)^2\partial_{y'}u^- - (k^-)^2\frac{2}{b}gu^- - (k^-)^2\frac{1}{b^2}g^2u^-. \qquad (2.3.1\text{m})$$

Furthermore,

$$J^+ = -\frac{1}{a}gT^+[u^+], \qquad (2.3.1\text{n})$$

$$J^- = \frac{1}{b}gT^-[u^-], \qquad (2.3.1\text{o})$$

and

$$\begin{aligned} Q \;=\; & \frac{1}{ab}\left\{(ab+ag-bg-g^2)(i\alpha\partial_{x'}g + i\beta)\phi(x') - ag\partial_{y'}u^+ \right. \\ & +(\partial_{x'}g)(b+g)(a-g)\partial_{x'}u^+ - (\partial_{x'}g)^2a(b+g)\partial_{y'}u^+ - bg\partial_{y''}u^- \\ & \left. -(\partial_{x'}g)(b+g)(a-g)\partial_{x'}u^- + (\partial_{x'}g)^2b(a-g)\partial_{y''}u^-\right\}. \end{aligned} \qquad (2.3.1\text{p})$$

2.3.2 Recursion by Boundary Perturbation

We shall now describe a boundary perturbation algorithm to solve the transformed system (2.3.1). If we let $g = \varepsilon f$ and f is sufficiently smooth, the transformed fields can be shown to be analytic. Hence, we can write

$$u^\pm(x,y;\varepsilon) = \sum_{n=0}^{\infty} u_n^\pm(x,y)\varepsilon^n.$$

Inserting the above into (2.3.1), it is straightforward, albeit tedious, to derive the following recursions for u_n:

$$\left(\partial_{x'}^2+\partial_{y'}^2\right)u_n^+(x',y')+(k^+)^2u_n^+(x',y')=F_n^+(x',y'),\qquad 0<y'<a, \tag{2.3.2a}$$

$$\partial_{y'}u_n^+(x',a)-T^+[u_n^+(x',a)]=J_n^+(x'), \tag{2.3.2b}$$

$$\left(\partial_{x'}^2+\partial_{y''}^2\right)u_n^-(x',y'')+(k^-)^2u_n^-(x',y'')=F_n^-(x',y''),\qquad -b<y''<0, \tag{2.3.2c}$$

$$\partial_{y''}u_n^-(x',-b)-T^-[u_n^-(x',-b)]=J_n^-(x'), \tag{2.3.2d}$$

$$u_n^+(x',0)-u_n^-(x',0)=\phi_n(x'), \tag{2.3.2e}$$

$$\partial_{y'}u_n^+(x',0)-\partial_{y''}u_n^-(x',0)=Q_n(x'). \tag{2.3.2f}$$

In these equations

$$F_n^\pm(x',y')=\partial_{x'}F_{n,x}^\pm(x',y')+\partial_{y'}F_{n,y}^\pm(x',y')+F_{n,h}^\pm(x',y'), \tag{2.3.2g}$$

where

$$F_{n,x}^+=\frac{2}{a}f\partial_{x'}u_{n-1}^+-\frac{1}{a^2}f^2\partial_{x'}u_{n-2}^++\frac{a-y'}{a}(\partial_{x'}f)\partial_{y'}u_{n-1}^+-\frac{a-y'}{a^2}f(\partial_{x'}f)\partial_{y'}u_{n-2}^+, \tag{2.3.2h}$$

$$F_{n,y}^+=\frac{a-y'}{a}(\partial_{x'}f)\partial_{x'}u_{n-1}^+-\frac{a-y'}{a^2}f(\partial_{x'}f)\partial_{x'}u_{n-2}^+-\frac{(a-y')^2}{a^2}(\partial_{x'}f)^2\partial_{y'}u_{n-2}^+, \tag{2.3.2i}$$

and

$$\begin{aligned}F_{n,h}^+=&-\frac{1}{a}(\partial_{x'}f)\partial_{x'}u_{n-1}^++\frac{1}{a^2}f(\partial_{x'}f)\partial_{x'}u_{n-2}^++\frac{a-y'}{a^2}(\partial_{x'}f)^2\partial_{y'}u_{n-2}^+\\&+(k^+)^2\frac{2}{a}fu_{n-1}^+-(k^+)^2\frac{1}{a^2}f^2u_{n-2}^+,\end{aligned} \tag{2.3.2j}$$

and

$$F_{n,x}^-=-\frac{2}{b}f\partial_{x'}u_{n-1}^--\frac{1}{b^2}f^2\partial_{x'}u_{n-2}^-+\frac{b+y'}{b}(\partial_{x'}f)\partial_{y'}u_{n-1}^-+\frac{b+y'}{b^2}f(\partial_{x'}f)\partial_{y'}u_{n-2}^-, \tag{2.3.2k}$$

$$F_{n,y}^-=\frac{b+y'}{b}(\partial_{x'}f)\partial_{x'}u_{n-1}^-+\frac{b+y'}{b^2}f(\partial_{x'}f)\partial_{x'}u_{n-2}^--\frac{(b+y')^2}{b^2}(\partial_{x'}f)^2\partial_{y'}u_{n-2}^-, \tag{2.3.2l}$$

and

$$\begin{aligned}F_{n,h}^-=&\frac{1}{b}(\partial_{x'}f)\partial_{x'}u_{n-1}^-+\frac{1}{b^2}f(\partial_{x'}f)\partial_{x'}u_{n-2}^--\frac{b+y'}{b^2}(\partial_{x'}f)^2\partial_{y'}u_{n-2}^-\\&-(k^-)^2\frac{2}{b}fu_{n-1}^--(k^-)^2\frac{1}{b^2}f^2u_{n-2}^-.\end{aligned} \tag{2.3.2m}$$

Furthermore,

$$J_n^+ = -\frac{1}{a} f T^+[u_{n-1}^+], \tag{2.3.2n}$$

$$J_n^- = \frac{1}{b} f T^-[u_{n-1}^-], \tag{2.3.2o}$$

$$\phi_n = (-1)^{n+1} \frac{(i\beta f)^n}{n!} e^{i\alpha x}, \tag{2.3.2p}$$

and

$$\begin{aligned} Q_n = \frac{1}{ab}\{ & -iab\beta\phi_n - iab\alpha\partial_{x'} f\phi_{n-1} - i\beta(a-b)f\phi_{n-1} - i\alpha(a-b)f\partial_{x'} f\phi_{n-2} + i\beta f^2\phi_{n-2} \\ & + i\alpha\partial_{x'} f f^2\phi_{n-3} - af\partial_{y'} u_{n-1}^+ + ab\partial_{x'} f\partial_{x'} u_{n-1}^+ + (a-b)f\partial_{x'} f\partial_{x'} u_{n-2}^+ \\ & - \partial_{x'} f f^2\partial_{x'} u_{n-3}^+ - ab(\partial_{x'} f)^2\partial_{y'} u_{n-2}^+ - a(\partial_{x'} f)^2 f\partial_{y'} u_{n-3}^+ - bf\partial_{y''} u_{n-1}^- - ab\partial_{x'} f\partial_{x'} u_{n-1}^- \\ & - (a-b)f\partial_{x'} f\partial_{x'} u_{n-2}^- + \partial_{x'} f f^2\partial_{x'} u_{n-3}^- + ab(\partial_{x'} f)^2\partial_{y''} u_{n-2}^- - bf(\partial_{x'} f)^2\partial_{y''} u_{n-3}^-\}. \end{aligned} \tag{2.3.2q}$$

If we write

$$u_n^\pm(x,y) = \sum_{p=-\infty}^{\infty} u_{n,p}^\pm(y) e^{i\alpha_p x},\ F_n^\pm(x,y) = \sum_{p=-\infty}^{\infty} F_{n,p}^\pm(y) e^{i\alpha_p x},\ J_n^\pm(x) = \sum_{p=-\infty}^{\infty} J_{n,p}^\pm e^{i\alpha_p x},$$

$$\phi_n(x) = \sum_{p=-\infty}^{\infty} \phi_{n,p} e^{i\alpha_p x},\ Q_n(x) = \sum_{p=-\infty}^{\infty} Q_{n,p}(y) e^{i\alpha_p x},$$

and insert into (2.3.2), then we obtain a sequence of equations for $u_{n,p}^\pm(y)$:

$$\begin{aligned} & \partial_{y'}^2 u_{n,p}^+(y') + ((k^+)^2 - \alpha_p^2) u_{n,p}^+(y') = F_{n,p}^+, && 0 < y' < a, \\ & \partial_{y''}^2 u_{n,p}^-(y'') + ((k^-)^2 - \alpha_p^2) u_{n,p}^-(y'') = F_{n,p}^-, && -b < y'' < 0, \\ & \partial_{y'} u_{n,p}^+(a) - i\beta_p^+ u_{n,p}^+(a) = J_{n,p}^+, \\ & \partial_{y''} u_{n,p}^-(-b) + i\beta_p^- u_{n,p}^-(-b) = J_{n,p}^-, \\ & u_{n,p}^+(0) - u_{n,p}^-(0) = \phi_{n,p}, \\ & \partial_{y'} u_{n,p}^+(0) - \partial_{y''} u_{n,p}^-(0) = Q_{n,p}. \end{aligned}$$

Due to the quasi-periodic boundary conditions we seek solutions of the form

$$u^\pm(x,y) = \sum_{n=0}^{\infty} \sum_{p=-\infty}^{\infty} u_{n,p}^\pm(y) e^{i\alpha_p x} \varepsilon^n,$$

resulting in the generic one–dimensional problem

$$\partial_y^2 u_{p,n}^+(y) + ((k^+)^2 - \tilde{\alpha}^2) u_{p,n}^+(y) = F_{p,n}^+(y), \qquad 0 < y < a, \tag{2.3.3a}$$

$$\partial_y^2 u_{p,n}^-(y) + ((k^-)^2 - \tilde{\alpha}^2) u_{p,n}^-(y) = F_{p,n}^-(y), \qquad -b < y < 0, \tag{2.3.3b}$$

$$\partial_y u_{p,n}^+(a) - i\beta^+ u_{p,n}^+(a) = J_{p,n}^+, \tag{2.3.3c}$$

$$\partial_y u_{p,n}^-(-b) + i\beta^- u_{p,n}^-(-b) = J_{p,n}^-, \tag{2.3.3d}$$

$$u_{p,n}^+(0) - u_{p,n}^-(0) = \phi_{p,n}, \tag{2.3.3e}$$

$$\partial_y u_{p,n}^+(0) - \partial_y u_{p,n}^-(0) = Q_{p,n}, \tag{2.3.3f}$$

where we have dropped the primes for convenience and denote

$$\beta^+ = \beta_p^+, \quad \beta^- = \beta_p^-, \quad \tilde{\alpha} = \alpha_p.$$

2.4 Legendre Galerkin Approximation

In this section we provide algorithm details of a Legendre Galerkin approach to approximate solutions of the two–point boundary value problem (2.3.3). The approximation of this problem is the final specification we must make in our TFE approach to the doubly layered scattering problem at hand.

2.4.1 Weak Formulation

Assume that $u^{*,+}(y)$ and $u^{*,-}(y)$ satisfy the following homogeneous version of (2.3.3):

$$\partial_y^2 u^{*,+} + ((k^+)^2 - \tilde{\alpha}^2) u^{*,+} = 0, \qquad 0 < y < a, \tag{2.4.1a}$$

$$\partial_y^2 u^{*,-} + ((k^-)^2 - \tilde{\alpha}^2) u^{*,-} = 0, \qquad -b < y < 0, \tag{2.4.1b}$$

$$\partial_y u^{*,+}(a) - i\beta^+ u^{*,+}(a) = J^+, \tag{2.4.1c}$$

$$\partial_y u^{*,-}(-b) + i\beta^- u^{*,-}(-b) = J^-, \tag{2.4.1d}$$

$$u^{*,+}(0) - u^{*,-}(0) = \phi, \tag{2.4.1e}$$

$$\partial_y u^{*,+}(0) - \partial_y u^{*,-}(0) = Q, \tag{2.4.1f}$$

where, for convenience, we have dropped the (n,p) subscripts. The functions

$$u^{*,+}(y) = Ae^{i\beta^+ y} + Be^{-i\beta^+ y}, \quad u^{*,-}(y) = Ce^{i\beta^- y} + De^{-i\beta^- y},$$

are solutions of (2.4.1a) & (2.4.1b), respectively, for any choices of the constants A, B, C, D. Substituting these forms into (2.4.1c)–(2.4.1f) we find A, B, C, D:

$$B = \frac{iJ^+ e^{i\beta^+ a}}{2\beta^+}, \quad C = -\frac{iJ^- e^{i\beta^- b}}{2\beta^-},$$
$$A = \frac{\beta^-(2C+\phi) + B(\beta^+ - \beta^-) - iQ}{\beta^- + \beta^+}, \quad D = \frac{\beta^+(2B-\phi) + C(\beta^- - \beta^+) - iQ}{\beta^- + \beta^+}.$$

Now consider the functions

$$\hat{u}^+(y) := u^+(y) - u^{*,+}(y), \quad \hat{u}^-(y) := u^-(y) - u^{*,-}(y),$$

where, $\hat{u}^+$ and $\hat{u}^-$ satisfy the version of (2.3.3) with homogeneous boundary conditions

$$\partial_y^2 \hat{u}^+(y) + ((k^+)^2 - \tilde{\alpha}^2)\hat{u}^+(y) = \hat{F}^+, \qquad 0 < y < a, \tag{2.4.2a}$$
$$\partial_y^2 \hat{u}^-(y) + ((k^-)^2 - \tilde{\alpha}^2)\hat{u}^-(y) = \hat{F}^-, \qquad -b < y < 0, \tag{2.4.2b}$$
$$\partial_y \hat{u}^+(a) - i\beta^+ \hat{u}^+(a) = 0, \tag{2.4.2c}$$
$$\partial_y \hat{u}^-(-b) + i\beta^- \hat{u}^-(-b) = 0, \tag{2.4.2d}$$
$$\hat{u}^+(0) - \hat{u}^-(0) = 0, \tag{2.4.2e}$$
$$\partial_y \hat{u}^+(0) - \partial_y \hat{u}^-(0) = 0. \tag{2.4.2f}$$

Setting

$$u(y) := \begin{cases} \hat{u}^+(y) & 0 < y < a \\ \hat{u}^-(y) & -b < y < 0 \end{cases}, \quad f(y) := \begin{cases} \hat{F}^+(y) & 0 < y < a \\ \hat{F}^-(y) & -b < y < 0 \end{cases},$$
$$k(y) := \begin{cases} (k^+)^2 - \tilde{\alpha}^2 & 0 < y < a \\ (k^-)^2 - \tilde{\alpha}^2 & -b < y < 0 \end{cases},$$

we find that u satisfies:

$$\partial_y^2 u(y) + k(y)^2 u(y) = f, \qquad -b < y < a, \tag{2.4.3a}$$

$$\partial_y u(a) - i\beta^+ u(a) = 0, \tag{2.4.3b}$$

$$\partial_y u(-b) + i\beta^- u(-b) = 0, \tag{2.4.3c}$$

$$u(0^+) - u(0^-) = 0, \tag{2.4.3d}$$

$$\partial_y u(0^+) - \partial_y u(0^-) = 0. \tag{2.4.3e}$$

Denoting the Sobolev space of complex functions:

$$H^1(-b, a) := \{u, \partial_y u \in L^2(-b, a)\},$$

we define the inner product on the interval $(-b, a)$

$$(u, v) := \int_{-b}^{a} u\bar{v}\, dy,$$

for any $u, v \in L^2(-b, a)$ where $\bar{v}$ is the complex conjugate of v. To simplify notation, we use from here the usual notation for spaces of real functions (e.g. H^1, P_N, etc) to denote spaces of complex functions.

With this notation the weak formulation for (2.4.3) is: Find $u \in H^1(-b, a)$ such that:

$$(k^2 u, \phi) - (\partial_y u, \partial_y \phi) = (f, \phi) - i\beta^+ u(a)\bar{\phi}(a) - i\beta^- u(-b)\bar{\phi}(-b), \qquad \forall \phi \in H^1(-b, a). \tag{2.4.4}$$

2.4.2 The Legendre-Galerkin Method

Let P_N be the polynomial space of degree at most N and define

$$X_{N,\beta,\gamma} := \{u \in C(-b, a) \mid u|_{(0,a)}, u|_{(-b,0)} \in P_N, (\partial_y u - i\beta u)(a) = (\partial_y u + i\gamma u)(-b) = 0\}.$$

Then our Legendre-Galerkin method is to find $u_N \in X_{N,\beta^+,\beta^-}$ such that

$$(k^2 u_N, \phi_N) - (\partial_y u_N, \partial_y \phi_N) = (\tilde{I}_N f, \phi_N) - i\beta^+ u_N(a)\bar{\phi}_N(a) - i\beta^- u_N(-b)\bar{\phi}_N(-b), \tag{2.4.5}$$

for all $\phi \in X_{N,\beta^+,\beta^-}$, where $\tilde{I}_N$ is the interpolation operator defined by $\tilde{I}_N f|_{(0,a)}$, $\tilde{I}_N f|_{(-b,0)} \in P_N$. Since every function in X_{N,β^+,β^-} is differentiable at everywhere except at zero, (2.4.5) is equivalent to

$$(k^2 u_N, \phi_N)+(\partial_y^2 u_N, \phi_N)_{I_1}+(\partial_y^2 u_N, \phi_N)_{I_2}+[\partial_y u_N(0^+)-\partial_y u_N(0^-)]\bar{\phi}_N(0) = (\tilde{I}_N f, \phi_N), \tag{2.4.6}$$

for all $\phi \in X_{N,\beta^+,\beta^-}$, where the subscripts I_1 and I_2 denote the corresponding integration domain $I_1 = (0, a)$ and $I_2 = (-b, 0)$.

Consider $\xi^+(y) = c_1 y + 1$ and $\xi^-(y) = c_2 y + 1$ such that

$$(\partial_y \xi^+ - i\beta^+ \xi^+)(a) = 0, \quad (\partial_y \xi^- + i\beta^- \xi^-)(-b) = 0.$$

It is not difficult to show that

$$c_1 = \frac{i\beta^+}{1 - i\beta^+ a}, \quad c_2 = \frac{-i\beta^-}{1 - i\beta^- b}.$$

If $L_j(y)$ is the Legendre polynomial of order j on $-1 < y < 1$, we define

$$\varphi_j(y) := (1+i)L_j\left(\frac{2y-a}{a}\right)+a_j L_{j+1}\left(\frac{2y-a}{a}\right)+b_j L_{j+2}\left(\frac{2y-a}{a}\right),\ j = 0, \ldots, N-2,$$

with the complex parameters a_j, b_j chosen such that φ_j satisfies the boundary conditions

$$(\partial_y \varphi_j - i\beta^+ \varphi_j)(a) = 0, \quad \varphi_j(0) = 0.$$

Similarly, we define

$$\psi_j(y) := (1+i)L_j\left(\frac{b+2y}{b}\right)+a'_j L_{j+1}\left(\frac{b+2y}{b}\right)+b'_j L_{j+2}\left(\frac{b+2y}{b}\right),\ j = 0, \ldots, N-2,$$

with a'_j, b'_j selected such that ψ_j satisfies the boundary conditions

$$(\partial_y \psi_j + i\beta^- \psi_j)(-b) = 0, \quad \psi_j(0) = 0.$$

If we let

$$\tilde{\phi}_j(y) := \begin{cases} \phi_j(y), & 0 < y < a, \\ 0, & -b < y < 0, \end{cases} \qquad j = 0, \ldots, N-2,$$

$$\tilde{\phi}_{N-1+j}(y) := \begin{cases} 0, & 0 < y < a, \\ \psi_j(y), & -b < y < 0, \end{cases} \qquad j = 0, \ldots, N-2,$$

$$\tilde{\phi}_{2N-2}(y) := \begin{cases} \xi^+(y), & 0 < y < a, \\ \xi^-(y), & -b < y < 0 \end{cases}.$$

Then, we have

$$X_{N,\beta^+,\beta^-} = \text{span}\{\tilde{\phi}_0, \tilde{\phi}_1, \ldots, \tilde{\phi}_{2N-2}\}.$$

We assume that the approximate solution has the form

$$u^N(y) := \sum_{j=0}^{2N-2} \hat{u}_j \tilde{\phi}_j(y), \tag{2.4.7}$$

and define

$$\begin{aligned} \hat{u} &= (\hat{u}_0, \ldots, \hat{u}_{N-2})^T \\ \hat{w} &= (\hat{u}_{N-1}, \ldots, \hat{u}_{2N-3})^T \\ \hat{f} &= (\hat{f}_0, \ldots, \hat{f}_{N-2})^T \\ \hat{g} &= (\hat{f}_{N-1}, \ldots, \hat{f}_{2N-3})^T \end{aligned}$$

where

$$\hat{f}_j := (\tilde{I}_N f, \tilde{\phi}_j), \quad j = 0, 1, \ldots, 2N-2.$$

We further define

$$\begin{aligned} s^1_{lj} &= (\partial_y^2 \tilde{\phi}_j, \tilde{\phi}_l)_{I_1}, \\ s^2_{lj} &= (\partial_y^2 \tilde{\phi}_{N-1+j}, \tilde{\phi}_{N-1+l})_{I_2}, \\ m^1_{lj} &= (\tilde{\phi}_j, \tilde{\phi}_l)_{I_1}, \\ m^2_{lj} &= (\tilde{\phi}_{N-1+j}, \tilde{\phi}_{N-1+l})_{I_2}, \end{aligned}$$

for $l, j = 0, 1, \ldots, N-2$. Additionally, we set

$$S_1 = (s^1_{lj}), \quad S_2 = (s^2_{lj}), \quad M_1 = (m^1_{lj}), \quad M_2 = (m^2_{lj}),$$
$$a_{12}(j) = (\partial_y^2 \tilde{\phi}_j + k^2 \tilde{\phi}_j, \tilde{\phi}_{2N-2})_{I_1} + \partial_y \tilde{\phi}_j(0^+),$$
$$b_{12}(j) = (\partial_y^2 \tilde{\phi}_{N-1+j} + k^2 \tilde{\phi}_{N-1+j}, \tilde{\phi}_{2N-2})_{I_2} - \partial_y \tilde{\phi}_{N-1+j}(0^-),$$
$$a_{21}(j) = (\partial_y^2 \tilde{\phi}_{2N-2} + k^2 \tilde{\phi}_{2N-2}, \tilde{\phi}_j)_{I_1},$$
$$b_{21}(j) = (\partial_y^2 \tilde{\phi}_{2N-2} + k^2 \tilde{\phi}_{2N-2}, \tilde{\phi}_{N-1+j})_{I_2},$$
$$a_{22}(j) = (k^2 \tilde{\phi}_{2N-2}, \tilde{\phi}_{2N-2}) + \partial_y \tilde{\phi}_{2N-2}(0^+) - \partial_y \tilde{\phi}_{2N-2}(0^-),$$
$$A_{11} = S_1 + (k^+)^2 M_1, \quad B_{11} = S_2 + (k^-)^2 M_2,$$

for $l, j = 0, 1, \ldots, N-2$. Upon insertion of (2.4.7) into (2.4.5) we find the following system of $2N-1$ equations:

$$\begin{pmatrix} A_{11} & 0 & a_{12} \\ 0 & B_{11} & b_{12} \\ a_{21}^T & b_{21}^T & a_{22} \end{pmatrix} \begin{pmatrix} \hat{u} \\ \hat{w} \\ \hat{u}_{2N-2} \end{pmatrix} = \begin{pmatrix} \hat{f} \\ \hat{g} \\ \hat{f}_{2N-2} \end{pmatrix}. \tag{2.4.8}$$

To solve this system of equations, we perform a simple block Gaussian elimination to get the following equation for $\hat{u}_{2N-2}$:

$$\left\{ a_{22} - \begin{pmatrix} a_{21}^T & b_{21}^T \end{pmatrix} \begin{pmatrix} A_{11} & 0 \\ 0 & B_{11} \end{pmatrix}^{-1} \begin{pmatrix} a_{12} \\ b_{12} \end{pmatrix} \right\} \hat{u}_{2N-2}$$
$$= \hat{f}_{2N-2} - \begin{pmatrix} a_{21}^T & b_{21}^T \end{pmatrix} \begin{pmatrix} A_{11} & 0 \\ 0 & B_{11} \end{pmatrix} \begin{pmatrix} \hat{f} \\ \hat{g} \end{pmatrix}.$$

Then, we can solve for $\hat{u}$ and $\hat{w}$ independently as follows:

$$A_{11}\hat{u} = \hat{f} - \hat{u}_{2N-2} \cdot (a_{12}),$$

and

$$B_{11}\hat{w} = \hat{g} - \hat{u}_{2N-2} \cdot b_{12}.$$

Due to the basis we chose, A_{11} and B_{11} are penta-diagonal symmetric matrices so that the above equations can be efficiently solved.

Finally, our numerical solution has the form

$$u^{N,N_x,N_y}(x,y) = \sum_{n=0}^{N} \sum_{p=-N_x/2}^{N_x/2-1} u_{n,p}^{N_y}(y) e^{i\alpha_p x} \varepsilon^n, \tag{2.4.9}$$

where

$$u_{n,p}^{N_y}(y) = u_{n,p}^{*}(y) + \sum_{j=0}^{2N_y-2} \hat{u}_{n,p,j} \tilde{\phi}_j(y)$$

with $u_{n,p}^{*}(y)$ from (2.4.1) and the $\hat{u}_{n,p,j}$ from the algorithm above using the Legendre Galerkin approximation.

Remark 2.4.1 Before leaving our description of the numerical procedure, we mention that there are a number of choices for summing the Taylor series which appear in (2.4.9). To avoid an avalanche of impenetrable notation we focus on the generic problem of approximating the analytic function

$$A(\varepsilon) = \sum_{n=0}^{\infty} A_n \varepsilon^n$$

by its truncated Taylor series

$$A^N(\varepsilon) := \sum_{n=0}^{N} A_n \varepsilon^n.$$

It is a classic result that if ε_0 is in the disk of convergence of $A(\varepsilon)$, say $\{|\varepsilon| < \rho\}$, $A^N(\varepsilon_0)$ will converge to $A(\varepsilon_0)$ exponentially fast as $N \to \infty$. However, it is possible for ε_0 to be a point of analyticity *outside* the disk of convergence of the Taylor series and for A^N to produce meaningless results. The classical numerical analytic continuation technique of Padé approximation [44] has been successfully brought to bear upon Boundary Perturbation Methods in the past (see, e.g., [8,18]) and we utilize this here as well. In short, Padé approximation seeks to simulate the truncated Taylor series A^N by the rational function

$$[L/M](\varepsilon) := \frac{a^L(\varepsilon)}{b^M(\varepsilon)} = \frac{\sum_{l=0}^{L} a_l \varepsilon^l}{1 + \sum_{m=1}^{M} b_m \varepsilon^m}$$

where $L + M = N$ and

$$[L/M](\varepsilon) = A^N(\varepsilon) + \mathcal{O}(\varepsilon^{L+M+1});$$

well-known formulas for the coefficients $\{a_l, b_m\}$ can be found in [44]. This approximant has the remarkable properties that, for a wide class of functions, not only is the convergence of $[L/M]$ to A at $\varepsilon = \varepsilon_0$ *faster* than that of A^N for $|\varepsilon_0| < \rho$, but also that $[L/M]$ may converge to A for points of analyticity ε_0 for which $|\varepsilon_0| > \rho$. We refer the interested reader to § 2.2 of Baker & Graves–Morris [44] and the insightful calculations of § 8.3 of Bender & Orszag [45] for a thorough discussion of the capabilities and limitations of Padé approximants.

2.5 Numerical Results & Discussion

We now present the results of numerical experiments which exhibit the stability and accuracy of our new algorithm. We use as a measure of convergence the widely–accepted "energy defect" [7–9, 46] and study the performance of our algorithm in assorted limits of both the physical and numerical parameters.

2.5.1 Energy Defect

To diagnose the convergence of our algorithm we appeal to the well–established energy conservation measure. We point out that *outside* the grooves, i.e. in the domain

$$\Omega_0 := \{y > |g|_{L^\infty}\} \bigcup \{y < -|g|_{L^\infty}\},$$

the solutions $u^\pm$ can be expressed via the Rayleigh expansions

$$u^+(x,y) = \sum_{p=-\infty}^{\infty} B_p^+ e^{i\alpha_p x + i\beta_p^+ y}, \quad u^-(x,y) = \sum_{p=-\infty}^{\infty} B_p^- e^{i\alpha_p x - i\beta_p^- y}. \tag{2.5.1}$$

In the case of real wavenumbers $k^\pm$ there is a principle of conservation of energy [46] for the TE mode which can be expressed as

$$\sum_{p\in U^+} \beta_p^+ \left|B_p^+\right|^2 + \sum_{p\in U^-} \beta_p^- \left|B_p^-\right|^2 = \beta_0^+.$$

Defining the energy

$$E^\pm(l) := \operatorname{Im}\left\{\frac{1}{L}\int_0^L \overline{u^\pm}(x,l)(\partial_y u^\pm(x,l))\, dx\right\}, \tag{2.5.2}$$

we have the following relationship:

Lemma 2.5.1 If $l_1 > |g|_{L^\infty}$ and $l_2 < -|g|_{L^\infty}$ then

$$E^+(l_1) - E^-(l_2) = \sum_{p \in U^+} \beta_p^+ \left|B_p^+\right|^2 + \sum_{p \in U^-} \beta_p^- \left|B_p^-\right|^2 = \beta_0^+.$$

Proof Simply substitute (3.4.11) into (3.4.13) and calculate the integral. ∎

Thus we can employ the "energy defect"

$$\delta := \left| 1 - \frac{E^+(l_1) - E^-(l_2)}{\beta_0^+} \right|,$$

to measure the error in our numerical approximation.

Before describing our results, we recall that k^+ and k^- are the wavenumbers in the upper and lower media, respectively, while α is the x–component of the incident radiation and ε measures the height/slope of our profile $y = g(x) = \varepsilon f(x)$ (which is always chosen $d = 2\pi$–periodic). In the first six examples in this section we have chosen $\alpha = 0$ (so that waves are normally incident) and selected the transparent boundaries at $y = a = 1$ and $y = -b = -1$. The numerical parameters are N_x (the number of Fourier modes in the x direction), N_y (the number of Legendre coefficients in the y direction), and N (the number of Taylor coefficients retained in the perturbation expansion).

In the recent work [42], a rigorous numerical error analysis of the TFE method was given for a single layer of dielectric material. We fully expect this analysis to apply directly to the doubly layered model at hand, and that our numerical approach will have very similar behavior, e.g., exponential convergence as N_x, N_y, and N are increased, and the need to increase all of these parameters as k^+ and k^- become large. However, we are also interested in two further questions which we address in the following numerical simulations:

1. As we increase ε so that the profile approaches the artificial boundaries, can we still obtain a reasonable approximation?

2. How does the difference between k^+ and k^- affect our results?

2.5.2 Numerical Results

We now perform a sequence of tests to study the convergence behavior of our algorithm.

1. Convergence study in perturbation order:

   To begin, we fix $d = 2\pi$, $\varepsilon = 0.1$, $N_x = 40$, $N_y = 80$, $f(x) = \cos(x)$, $a = 1$, $b = -1$, and vary $N = 0, \ldots, 55$ for five choices of the wavenumbers $k^{\pm}$:

$$(k^+, k^-) = (2.5, 1.25), \quad (k^+, k^-) = (12.5, 6.25), \quad (k^+, k^-) = (25.5, 12.75),$$
$$(k^+, k^-) = (51.5, 25.75), \quad (k^+, k^-) = (102.5, 51.25). \tag{2.5.3}$$

   The results are displayed in Figure 2.2. Clearly, as anticipated, we notice ex-

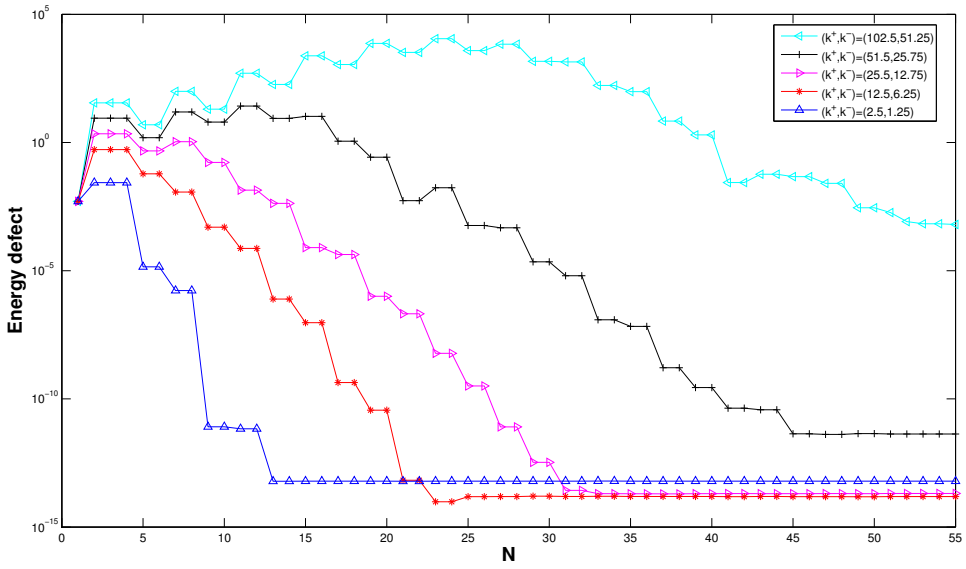


Figure 2.2. Energy defect versus perturbation order N.

   ponential convergence as N is refined. We point out that larger values of the wavenumbers require much larger choices for N, and, furthermore, as we fixed the x and y discretizations at $N_x = 40$ and $N_y = 80$, respectively, the case

$(k^+, k^-) = (102.5, 51.25)$ is under-resolved and we can only achieve an error of 10^{-3}.

2. Convergence study in vertical discretization:

   We now fix $d = 2\pi$, $\varepsilon = 0.1$, $N_x = 20$, $N = 20$, $f(x) = \cos(x)$, $a = 1$, $b = -1$, and vary $N_y = 1, \ldots, 40$ for the five choices of $k^\pm$ in (2.5.3). We display the results in Figure 2.3. Once again, we notice exponential convergence as N_y is

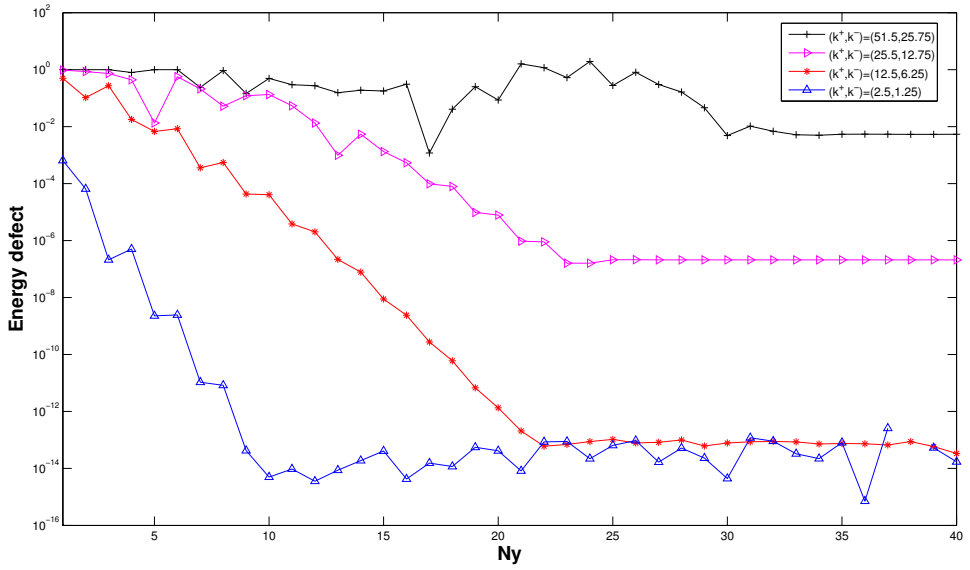


Figure 2.3. Energy defect versus vertical discretization N_y.

refined, and larger values of the wavenumbers require much larger choices for N_y. Once again, the calculation is under-resolved at $(k^+, k^-) = (102.5, 51.25)$ so that we can only realize an error of 10^{-2} and is thus omitted.

3. Convergence study in horizontal discretization:

   We fix $d = 2\pi$, $\varepsilon = 0.1$, $N_y = 80$, $N = 20$, $f(x) = \cos(x)$, $a = 1$, $b = -1$, and vary $N_x = 1, \ldots, 40$ for the first four choices of $k^\pm$ in (2.5.3). We display the results in Figure 2.4. Exponential convergence is once again observed, though

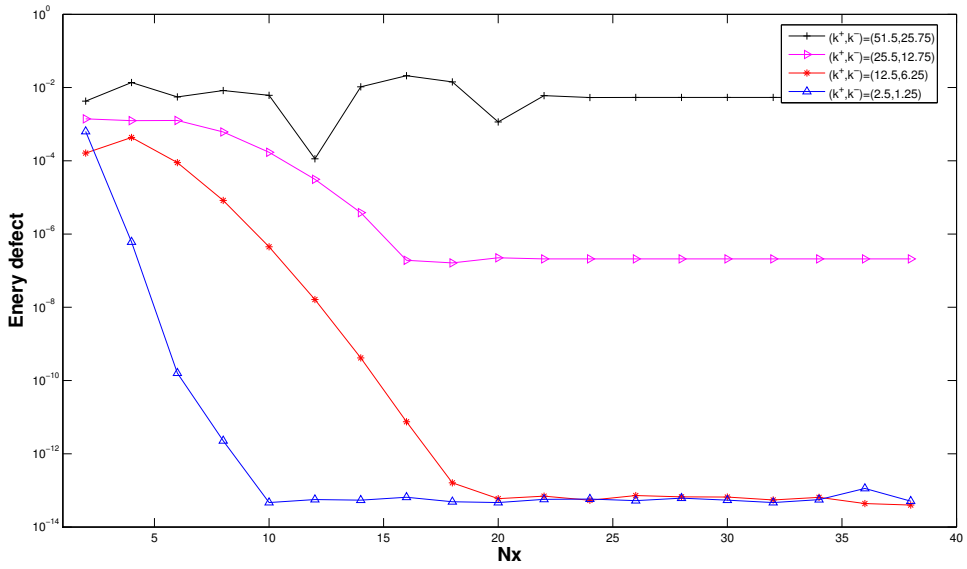


Figure 2.4. Energy defect versus horizontal discretization N_x.

for the two larger choices of wavenumber the under-resolution is particularly strong in this calculation.

4. Convergence study for deformations near the artificial boundary:

   We now investigate the behavior of our algorithm as the sharp interface and artificial boundary are brought close together. This can, of course, be achieved either by increasing ε, decreasing a (or b) or a combination of both. To fix upon an example we set $d = 2\pi$, $f(x) = \cos(x)$, $N_x = 20$, and $N_y = 40$. We investigate five configurations:

$$\begin{aligned}
(k^+, k^-, N) &= (2.5, 1.25, 30), \\
(k^+, k^-, N) &= (2.5, 1.25, 50), \\
(k^+, k^-, N) &= (12.5, 6.25, 30), \\
(k^+, k^-, N) &= (12.5, 6.25, 80), \\
(k^+, k^-, N) &= (12.5, 6.25, 200),
\end{aligned}$$

and, letting $\varepsilon = 0.1, 0.2, \ldots, 0.9$, we display the results in Figure 2.5. First, from

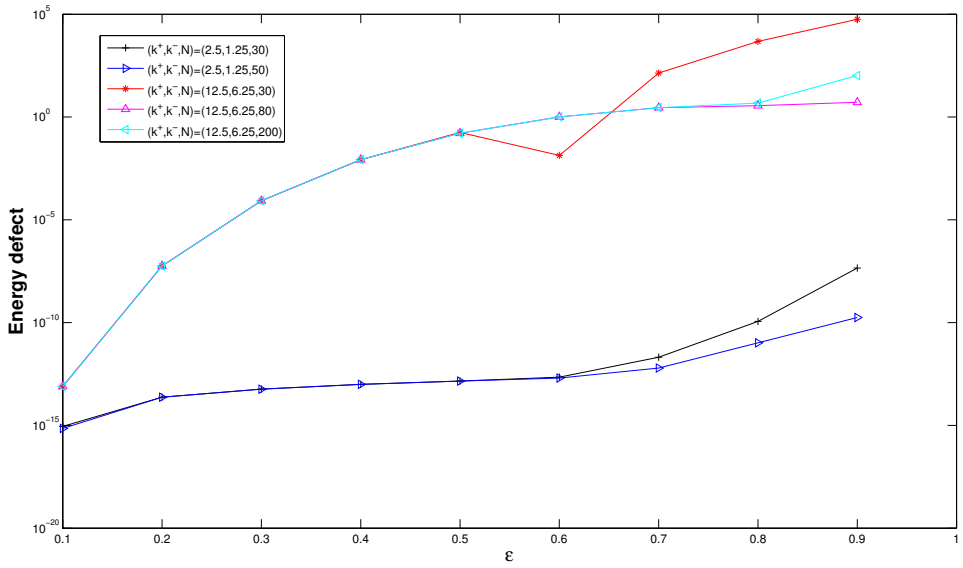


Figure 2.5. Energy defect versus ε.

the previous sections, we observe that parameters $(N_x, N_y, N) = (20, 40, 30)$ are sufficiently large to obtain a high accuracy approximation both for the cases $(k^+, k^-) = (2.5, 1.25)$ and $(k^+, k^-) = (12.5, 5.25)$. Figure 2.5 shows that when we let the height of the profile approach the artificial boundaries, the error is only determined by the parameters (k^+, k^-). To achieve the same relative error, for small (k^+, k^-) one can allow the artificial boundaries to be located quite close to the profile. Here Padé approximation (see Remark 2.4.1) was used to access this region of extended analyticity so that configurations which are large deformations of the base geometry can be simulated [44].

5. Convergence study as wavenumber is varied:

   We investigate the effects of varying the ratio of the wavenumber parameters k^-/k^+ in our numerical scheme. We fix $d = 2\pi$, $f(x) = \cos(x)$, $N_x = 40$, $N_y = 80$, $N = 30$, $\varepsilon = 0.1$, $k^+ = 2.5$, and vary k^-. From Table 2.1, we can see

Table 2.1
Energy defect versus wavenumber ratio k^-/k^+

k^-/k^+	Energy Defect	k^-/k^+	Energy Defect
0/20	$1.068656274583191 \times 10^{-12}$	10/20	$5.115907697472721 \times 10^{-14}$
1/20	$5.464073637995171 \times 10^{-13}$	11/20	$2.842170943040401 \times 10^{-15}$
2/20	$6.931344387339777 \times 10^{-13}$	12/20	$8.792966355031240 \times 10^{-14}$
3/20	$7.194245199571014 \times 10^{-13}$	13/20	$1.820765760385257 \times 10^{-13}$
4/20	$1.673328142715036 \times 10^{-13}$	14/20	$2.952305067083216 \times 10^{-13}$
5/20	$3.371525281181676 \times 10^{-13}$	15/20	$1.813660333027656 \times 10^{-13}$
6/20	$8.196110456992755 \times 10^{-13}$	16/20	$3.307576434963266 \times 10^{-13}$
7/20	$1.231015289704374 \times 10^{-13}$	17/20	$3.364419853824074 \times 10^{-13}$
8/20	$8.864020628607250 \times 10^{-14}$	18/20	$4.991562718714704 \times 10^{-14}$
9/20	$1.353583911622991 \times 10^{-13}$	19/20	$7.744915819785092 \times 10^{-14}$

that the difference between k^+ and k^- has almost no effect on the error if we choose parameters N_x, N_y, and N large enough.

6. Convergence study as energy defect is fixed:

   We fix $d = 2\pi$, $f(x) = \cos(x)$, $\varepsilon = 0.1$, and aim to find the smallest set of resolution parameters (N, N_x, N_y) for a range of wavenumber pairs (k^+, k^-) such that an energy defect smaller than 10^{-6} is achieved. The result is listed in Table 2.2. We observe that only a moderately number of modes/iterations, which grow linearly as the wavenumber increases, are needed to obtain an accuracy of 10^{-6}.

7. Convergence study as incident wave angle is varied:

   For our final study, we investigate the effects of varying the incident wave angle parameter $\cos\theta = \alpha/k^+$ in our numerical scheme. When $\alpha = 0$, it means the wave is normally incident ($\theta = \frac{\pi}{2}$). We fix $d = 2\pi$, $f(x) = \cos(x)$, $N_x = 10$, $N_y = 15$, $N = 12$, $\varepsilon = 0.1$, $k^+ = 12.5$, $k^- = 6.25$ and vary α. We observe from

Table 2.2
Smallest (N, N_x, N_y) for (k^+, k^-) to achieve an error of 10^{-6}.

(k^+, k^-)	N	N_x	N_y
(2.5,1.25)	7	6	7
(12.5,6.25)	12	10	15
(25.5, 12,75)	20	18	24
(51.5,25.75)	26	26	42
(85.5,42.75)	50	38	63
(105.5,52.75)	60	60	74

Table 2.2 that $N_x = 10$, $N_y = 15$, $N = 12$ are the smallest numbers to achieve an accuracy of 10^{-6} when $(k^+, k^-, \alpha) = (12.5, 6.25, 0)$. We observe from Table

Table 2.3
Energy defect versus incident wave angle α/k^+

α/k^+	Energy Defect	α/k^+	Energy Defect
0/10	$5.186341816312279 \times 10^{-6}$	5/10	$2.292137464521894 \times 10^{-5}$
1/10	$1.432592737851698 \times 10^{-5}$	6/10	$7.991322704206624 \times 10^{-6}$
2/10	$9.598113105472218 \times 10^{-6}$	7/10	$6.234791904884309 \times 10^{-6}$
3/10	$7.528997959674400 \times 10^{-6}$	8/10	$3.751204664833842 \times 10^{-6}$
4/10	$1.334178326434116 \times 10^{-5}$	9/10	$9.947091869799839 \times 10^{-7}$

2.3 that different α and k^+ have very little effect on the accuracy for a fixed set of parameters N_x, N_y, and N.

2.6 Conclusion

We constructed and implemented a Boundary Perturbation Method for the scattering of electromagnetic waves by doubly layered periodic dielectric media. The method is based on three essential steps: (i) a domain flattening through a change of variable; (ii) a recursion by boundary perturbation; and (iii) an efficient and accurate Legendre-Galerkin method for solving the one–dimensional Helmholtz equation with piecewise constant wavenumbers. The resulting algorithm is shown to be very efficient and stable for a range of small to moderate wavenumbers. On the other hand, our method is not specially designed for the technologically important high–frequency case (reflected in our equations with large values of k). While not beyond the scope of our method, such a simulation would require a very fine discretization of the problem domain resulting in an enormous count of degrees of freedom.

While we have only considered the two–dimensional doubly layered dielectric media, it is expected that the method can be extended to two–dimensional multi–layered periodic media, as well as three–dimensional Maxwell's equations with doubly periodic multi–layered media.

CHAPTER 3. A SPECTRAL ELEMENT METHOD WITH TRANSPARENT BOUNDARY CONDITIONS FOR ACOUSTIC TIME-HARMONIC SCATTERING IN PERIODIC DOUBLE-LAYER STRUCTURES

In this chapter, we present a spectral element method in combination with a transparent boundary condition for simulating acoustic scattering problems with nonreflecting waves on the truncated computational domain. We consider time-harmonic scattering problem represented by the Helmholtz equation. We approximate solution by a tensor product basis of the one-dimensional Legendre-Lagrange interpolation polynomials based on the Gauss-Lobatto-Legendre grids. Geometric structures are accurately represented by body-fitted quadrilateral elements. We use a GMRES iteration technique to solve the resulting system of the equations. We demonstrate computational results for the geometries with smooth flat structures, smooth curved structures, and non smooth interface structures. We validate our results with convergence studies in comparison with exact solutions, FTE solutions, and the energy defect measure. Our results show exponential convergence with respect to the approximation order.

3.1 Introduction

In Chapter 2, we point out that the scattering of electromagnetic or acoustic waves by a periodic structure plays an important role in many engineering problems. We also develop a high–order numerical method, based on the TFE approach, to solve the Helmholtz equation which can be used to simulating the interaction of time–harmonic electromagnetic waves incident upon a periodic doubly layered dielectric material with a sharply defined, irregular interface. However, for the non smooth structure, the convergence of TFE approach is highly reduced. Therefore, in this work, we shall

describe a alternative high-order Spectral Element Method (SEM) in combination with a transparent boundary condition for the numerical simulation of scattering returns from an irregularly shaped, periodic, doubly layered medium, which combines the geometry flexibility of the finite element method with the accuracy of spectral method. Compared with the TFE approach, however, the SEM approach needs to solve a two-dimensional problem, which shall result in a large, sparse, ill-conditioned indefinite matrix system, and make it computationally more expensive. While we focus upon periodic structures given the large number of engineering applications they inspire, this choice does simplify our numerical approach (e.g., we may use the Discrete Fourier Transform to approximate Fourier coefficients). However, we note that this simplification is also realized for many competing methods as well. One contribution of this work is that we propose a new approach to compute the discrete Fourier expansion based on the relation between spherical Bessel functions and Legendre functions, which can be more efficient than the other approach.

Of particular importance to the current project, this method was generalized to precisely the case we treat here (a doubly layered, two–dimensional material) by two of the authors [19]. We use this algorithm as a test for the numerical approach we advocate in this chapter and with it we can illuminate the strengths and shortcomings of this new SEM approach.

This Chapter is organized as follows. In Section § 3.2, we present the governing equations. In Section § 3.2.3 – § 3.3, we discuss the variational formulation and discretizations. Section § 3.3 demonstrates computational results and their validation provided with convergence studies. Section § 3.4 gives the conclusion.

3.2 Governing Equations

Consider a monochromatic plane wave with time dependent $e^{-\mathbf{i}wt}$ propagating in x-y plaine. Time-harmonic acoustic scattering can be described by the Helmholtz equation:

$$\Delta U + k^2 U = 0, \tag{3.2.1}$$

where $k = \frac{w}{c}$ is the wavenumber with the relation to the angular frequency w and the speed of sound c. The total field

$$U(x, y) = e^{\mathbf{i}(\alpha x + \beta y)}. \tag{3.2.2}$$

Acoustic scattering waves in layered periodic structures whose domain is defined by

$$\Omega^+ := \{y > g(x)\}, \quad \Omega^- := \{y < g(x)\}, \tag{3.2.3}$$

where $y = g(x)$ represents the shape of the d–periodic interface as shown in Figure 3.1. We study the scattering of acoustic plane waves from impedance gratings in layered

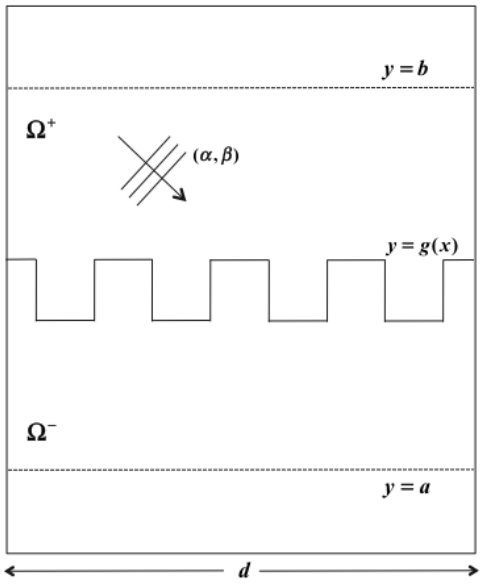


Figure 3.1. Geometric illustration of the problem

media. The acoustic time-harmonic planewave at the frequency ω can be represented by

$$U(x, y, t) = u(x, y)e^{-\mathbf{i}\omega t}. \tag{3.2.4}$$

With the periodicity of the grating enforces the quasi-periodicity of the fields, we have

$$u(x + d, y) = e^{\mathbf{i}\alpha d}u(x, y), \tag{3.2.5}$$

and the scattring waves must be outgoing. Consider the acoustic time-harmonic scattering described by the Helmholtz equation as

$$\Delta u + k^2 u = 0, \tag{3.2.6}$$

where $k = \omega/c$. Thus, the governing equations we consider are

$$\Delta u + k^2 u = 0 \tag{3.2.7}$$

$$\mathrm{OWC}[u] = 0, \qquad y \to \pm\infty, \tag{3.2.8}$$

$$u(x + d, y) = e^{\mathbf{i}\alpha d}u(x, y), \tag{3.2.9}$$

and we make the "outgoing wave condition" (OWC) operators more precise presently.

3.2.1 Transparent Boundary Conditions

For the far-field boundary conditions consider the hyperplanes $\{y = a\}$, $\{y = b\}$ where $|a|, |b| > |g|_{L^\infty}$. The augmented system of governing equations on R^2 and $\Omega = [0, L] \times [a, b]$.

$$\Delta u + k^2 u = 0, \qquad \text{on } \Omega \tag{3.2.10}$$

$$\partial_y u = \partial_y v, \qquad \text{on } \Gamma \tag{3.2.11}$$

$$u = v, \qquad \text{on } \Gamma \tag{3.2.12}$$

$$\Delta v + k^2 v = 0, \qquad \text{on } R^2 - \Omega \tag{3.2.13}$$

$$\mathrm{OWC}[v] = 0, \qquad y \to \pm\infty, \tag{3.2.14}$$

$$u(x + d, y) = e^{\mathbf{i}\alpha d} u(x, y), \tag{3.2.15}$$

$$v(x + d, y) = e^{i\alpha d} v(x, y), \tag{3.2.16}$$

are equivalent to (3.2.7). To make the far-field boundary condition more precise we note that solutions of (3.2.13)–(3.2.14) are

$$v(x, y) = \sum_{p=-\infty}^{\infty} \hat{\psi}_p e^{\mathbf{i}\alpha_p x + \mathbf{i}\beta_p (y-b)}, \tag{3.2.17}$$

where

$$\alpha_p := \alpha + (2\pi/d)p, \quad \beta_p := \begin{cases} \sqrt{k^2 - \alpha_p^2} & p \in K \\ \mathbf{i}\sqrt{\alpha_p^2 - k^2} & p \notin K \end{cases}, \tag{3.2.18}$$

$$K := \left\{ p \in \mathbf{Z} \mid k^2 - \alpha_p^2 > 0 \right\}, \tag{3.2.19}$$

$\mathbf{Z}$ are the integers, and

$$\psi(x) := u(x, a) = \sum_{p=-\infty}^{\infty} \hat{\psi}_p e^{\mathbf{i}\alpha_p x}. \tag{3.2.20}$$

Similarly, solutions of (2.2.4i)–(2.2.4j) are

$$v(x, y) = \sum_{p=-\infty}^{\infty} \hat{\zeta}_p e^{\mathbf{i}\alpha_p x - \mathbf{i}\beta_p (y-a)}, \tag{3.2.21}$$

where

$$\zeta(x) := u(x,a) = \sum_{p=-\infty}^{\infty} \hat{\zeta}_p e^{i\alpha_p x}. \tag{3.2.22}$$

To close the set of equations for u we simply need to produce $\partial_y v$ in (3.2.11):

$$\partial_y v(x,b) = \sum_{p=-\infty}^{\infty} \mathbf{i}\beta_p \hat{\psi}_p e^{\mathbf{i}\alpha_p x} =: T^+[\psi] = T^+[u(x,b)]. \tag{3.2.23}$$

A similar analysis at $y = a$ yields an operator

$$T^-[\zeta] := \sum_{p=-\infty}^{\infty} (-\mathbf{i}\beta_p)\hat{\zeta}_p e^{\mathbf{i}\alpha_p x}, \tag{3.2.24}$$

and the system (2.2.4) can be *equivalently* restated as

$$\Delta u + k^2 u = 0, \qquad g(x) < y < b, \tag{3.2.25}$$

$$\partial_y u - T^+[u] = 0, \qquad y = b, \tag{3.2.26}$$

$$\partial_y u - T^-[u] = 0, \qquad y = a, \tag{3.2.27}$$

$$u(x+d,y) = e^{\mathbf{i}\alpha d} u(x,y). \tag{3.2.28}$$

3.2.2 Quasi-Periodic Formulation

In our computation, we consider arbitray incident waves for $\alpha \neq 0$. To keep the periodicity, we introduce a new variable w. Let $u(x,y) = e^{-\mathbf{i}\alpha x} U(x,y)$, then $u(x+d,y) = e^{-\mathbf{i}\alpha(x+d)} U(x+d,y) = e^{-\mathbf{i}\alpha(x+d)+d} U(x,y) = e^{-\mathbf{i}\alpha x} U(x,y) = u(x,y)$, so we have $u(x,y)$ be periodic functions. If we only consider the equation for total field $U(x,y) = U^{\text{inc}} + U^{\text{scat}}$, then the total field will be continuous across the interface and satisfy the following equations:

$$\Delta U + k^2 U = 0, \qquad \text{on } \Omega, \tag{3.2.29a}$$

$$\partial_y U - T^+[U] = \partial_y U^{\text{inc}} - T^+[U^{\text{inc}}], \qquad \text{on } \Gamma^+, \tag{3.2.29b}$$

$$\partial_y U - T^-[U] = 0, \qquad \text{on } \Gamma^-, \tag{3.2.29c}$$

$$U(x+d,y) = e^{\mathbf{i}\alpha d} U(x,y). \tag{3.2.29d}$$

Furthermore, if we define

$$u(x,y) = e^{-\mathbf{i}\alpha x}U(x,y) \;\text{ and }\; u^{\text{inc}}(x,y) = e^{-\mathbf{i}\alpha x}U^{\text{inc}}(x,y), \tag{3.2.30}$$

then the new variable is periodic and satisfy the following equations:

$$\Delta u + (k^2-\alpha^2)u + 2i\alpha\frac{\partial u}{\partial x} = 0, \qquad \text{on } \Omega, \tag{3.2.31a}$$

$$\partial_y u - T_0^+[u] = \rho(x) \qquad \text{on } \Gamma^+ \tag{3.2.31b}$$

$$\partial_y u - T_0^-[u] = 0, \qquad \text{on } \Gamma^-, \tag{3.2.31c}$$

$$u(x+d,y) = u(x,y). \tag{3.2.31d}$$

where $\rho(x) = \partial_y u^{\text{inc}} - T_0^+[u^{\text{inc}}]$ and

$$T_0^+[u(x,a)] = \sum_{p=-\infty}^{\infty}(\mathbf{i}\beta_p)\hat{u}_p e^{\mathbf{i}(\alpha_p-\alpha)x} = \sum_{p=-\infty}^{\infty}(\mathbf{i}\beta_p)\hat{u}_p e^{\mathbf{i}(2\pi/d*p)x} \tag{3.2.32}$$

3.2.3 Variational Formulation

In our computation, we solve the model problem (3.2.31) with the following variational formulation:

$$a(u,\upsilon) = \langle\rho,\upsilon\rangle_{\Gamma_a} \qquad \text{for all } \upsilon \in H^1_{\Gamma_g}(\Omega), \tag{3.2.33}$$

where the sesquilinear form is defined as

$$a(u,\upsilon) := \int_\Omega(\nabla u\cdot\nabla\overline{\upsilon} - (k^2-\alpha^2)u\overline{\upsilon} + 2\mathbf{i}\alpha\frac{\partial u}{\partial x}\overline{\upsilon})d\Omega - \int_{\Gamma_a}T_0^+[u]\overline{\upsilon}d\Gamma - \int_{\Gamma_b}T_0^-[u]\overline{\upsilon}d\Gamma, \tag{3.2.34}$$

and the linear functionals are given by

$$\langle\rho,\upsilon\rangle_{\Gamma_a} := \int_{\Gamma_a}\rho\overline{\upsilon}d\Gamma, \tag{3.2.35}$$

and boundary specified by

$$\Gamma_g := \{(x,y) \mid y = g(x), x\in[0,d]\}, \quad \Gamma_a := \{(x,y) \mid y = a, x\in[0,d]\}. \tag{3.2.36}$$

If we denote

$$a_1(u,\upsilon) := \int_\Omega(\nabla u\cdot\nabla\overline{\upsilon} - (k^2-\alpha^2)u\overline{\upsilon})d\Omega \tag{3.2.37}$$

$$a_2(u,\upsilon) := \int_\Omega 2\mathbf{i}\alpha\frac{\partial u}{\partial x}\overline{\upsilon}d\Omega \tag{3.2.38}$$

and

$$a_3(u,\upsilon) := \int_\Gamma T_0^+[u]\overline{\upsilon}d\Gamma + \int_\Gamma T_0^-[u]\overline{\upsilon}d\Gamma, \tag{3.2.39}$$

If $\alpha = 0$, then $A_2 = 0$ obviously, $A_1 = A_1^* = A_1^T$ as $a_1(u,\upsilon) = \overline{a_1(\upsilon,u)}$ and A_1 is real.

$$(T^\pm[u],\bar{v})_\Gamma := \sum_{p=-\infty}^{\infty} \pm\mathbf{i}\beta_p^\pm\hat{u}_p(e^{\mathbf{i}\alpha_p x},\bar{v}) = \sum_{p=-\infty}^{\infty} \pm\mathbf{i}\beta_p^\pm\hat{u}_p\overline{(e^{-\mathbf{i}\alpha_p x},v)} = \sum_{p=-\infty}^{\infty} \pm\mathbf{i}\beta_p^\pm\hat{u}_p\overline{\hat{v}_p} \tag{3.2.40}$$

and

$$(T^\pm[v],\bar{u})_\Gamma = \sum_{p=-\infty}^{\infty} \pm\mathbf{i}\beta_p^\pm\hat{v}_p\overline{\hat{u}_p} \tag{3.2.41}$$

3.3 Spectral Element Discretization

We denote our computational domain in two dimensions as $\Omega = \cup_{e=1}^E\Omega^e$, where Ω^e represents nonoverlapping body-conforming quadrilateral elements. Let us define a finite-dimensional approximation space $V_N \subset H^1(\Omega)$, such that

$$V_N = \text{span}\{\psi_{ij}(\xi,\eta)\}_{i,j=0}^N.$$

We map each physical coordinate $(x,y) \in \Omega^e$ onto the reference domain $(\xi,\eta) \in I = [-1,1]^2$ through the Gordon–Hall mapping [47]. With this choice of approximation space, we consider a local approximate solution $u^e(x,y) \in V_N$, or simply u^e, that has the representation

$$u^e(x,y) = \sum_{i,j=0}^N u_{ij}^e\psi_{ij}(\xi,\eta), \tag{3.3.1}$$

where the basis coefficients u_{ij}^e are the nodal values $u^e(x_i,y_j)$ on Ω^e and $\psi_{ij}(\xi,\eta) = \ell_i(\xi)\ell_j(\eta)$, or simply ψ_{ij}, are the tensor product basis of the one-dimensional Nth-order Legendre-Lagrange interpolation polynomials defined as

$$\ell_i(\xi) = [N(N+1)^{-1}(1-\xi^2)L_N'(\xi)]/[(\xi-\xi_i)L_N(\xi_i)] \quad \text{for } \xi \in [-1,1], \tag{3.3.2}$$

based on the Gauss-Lobatto-Legendre (GLL) quadrature nodes ξ_i with the derivative of the Nth-order Legendre polynomial L'_N.

Let us denote our numerical solution on Ω as $\mathbf{u}$ with vector representations as

$$\mathbf{u} := (u_1, u_2, ..., u_{\hat{l}}, ..., u_n) := (u^1, u^2, ..., u^e, ..., u^E)^T, \tag{3.3.3}$$

$$u^e := (u^e_1, u^e_2, ..., u^e_l, ..., u^e_{(N+1)(N+1)})^T := (u^e_{00}, u^e_{10}, ..., u^e_{ij}, ..., u^e_{NN})^T, \tag{3.3.4}$$

where $n = E(N+1)^2$ is the total number of basis coefficients, and $\hat{l} = 1 + i + j(N+1) + (e-1)(N+1)^2$ and $l = 1 + i + j(N+1)$ translate the two-index coefficient representation into a vector form, with the leading index i. In Figure 3.2, we show a mesh with two elements $E = 2$ including the GLL grids for $N = 3$ on $\Omega = \Omega^1 \cup \Omega^2$. Figure 3.2(a) illustrates the local ordering of the solution vector $\mathbf{u}$ using the two-index expression, based on the unassembled representation including the coincident grids, $u^1_{3i} = u^2_{0i}$ $(i = 0, ..., 3)$, redundantly. Here we introduce the solution vector based on the global ordering, the assembled representation using only the distinct nodes as illustrated in Figure 3.2(b),

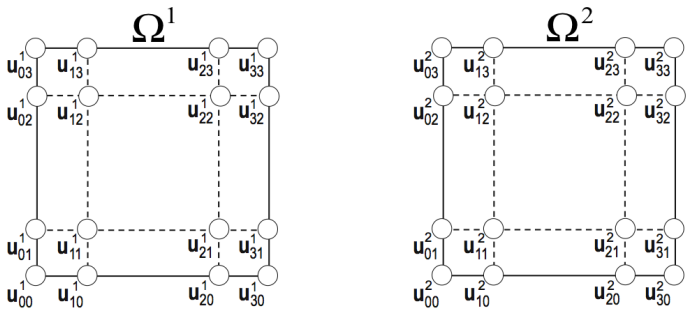
$$\underline{\mathbf{u}} = (\underline{u}_1, \underline{u}_2, ..., \underline{u}_{\bar{n}})^T, \tag{3.3.5}$$

where the size $(\bar{n} < n)$ of the solution vector $\underline{\mathbf{u}}$ in global ordering is reduced after eliminating the redundancy from the coincident grids.

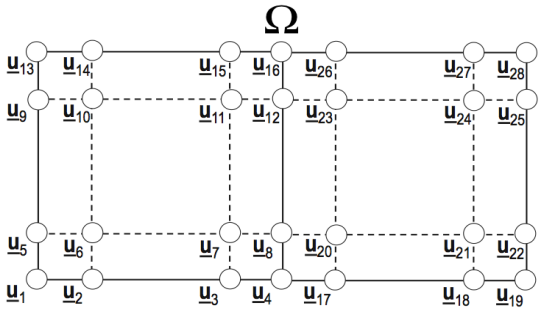
3.3.1 Stiffness Matrices

To obtain the stiffness matrix, we consider the inner product in Eq. (3.2.34):

$$A(u, v) = \int_\Omega \nabla u \cdot \nabla \overline{v} d\Omega = \int_\Omega \left(\frac{\partial u}{\partial x}\frac{\partial \overline{v}}{\partial x} + \frac{\partial u}{\partial y}\frac{\partial \overline{v}}{\partial y} \right) d\Omega, \tag{3.3.6}$$



(a) Local Numbering (Unassembled Representation)



(b) Global Numbering (Assembled Representation)

Figure 3.2. Illustration of the solution vector in the local numbering and the global numbering on an example mesh with $(E, N) = (2, 3)$: $\Omega = \Omega^1 \cup \Omega^2$ and the GLL nodes ($\circ$).

where the partial derivatives are expressed by the chain rule for $x = x(\xi, \eta)$ and $y = y(\xi, \eta)$ on Ω^e:

$$\begin{aligned}
\frac{\partial u}{\partial x}\frac{\partial \overline{\upsilon}}{\partial x} &= \left(\frac{\partial u}{\partial \xi}\frac{\partial \xi}{\partial x} + \frac{\partial u}{\partial \eta}\frac{\partial \eta}{\partial x}\right)\left(\frac{\partial \overline{\upsilon}}{\partial \xi}\frac{\partial \xi}{\partial x} + \frac{\partial \overline{\upsilon}}{\partial \eta}\frac{\partial \eta}{\partial x}\right) \\
&= \frac{\partial u}{\partial \xi}\frac{\partial \overline{\upsilon}}{\partial \xi}\left(\frac{\partial \xi}{\partial x}\frac{\partial \xi}{\partial x}\right) + \frac{\partial u}{\partial \eta}\frac{\partial \overline{\upsilon}}{\partial \eta}\left(\frac{\partial \eta}{\partial x}\frac{\partial \eta}{\partial x}\right) + \frac{\partial u}{\partial \xi}\frac{\partial \overline{\upsilon}}{\partial \eta}\left(\frac{\partial \xi}{\partial x}\frac{\partial \eta}{\partial x}\right) + \frac{\partial u}{\partial \eta}\frac{\partial \overline{\upsilon}}{\partial \xi}\left(\frac{\partial \xi}{\partial x}\frac{\partial \eta}{\partial x}\right) \\
&= \frac{\partial u}{\partial \xi}\frac{\partial \overline{\upsilon}}{\partial \xi}\mathcal{G}^{\xi\xi}_{xx} + \frac{\partial u}{\partial \eta}\frac{\partial \overline{\upsilon}}{\partial \eta}\mathcal{G}^{\eta\eta}_{xx} + \frac{\partial u}{\partial \xi}\frac{\partial \overline{\upsilon}}{\partial \eta}\mathcal{G}^{\xi\eta}_{xx} + \frac{\partial u}{\partial \eta}\frac{\partial \overline{\upsilon}}{\partial \xi}\mathcal{G}^{\xi\eta}_{xx}, \qquad (3.3.7) \\
\frac{\partial u}{\partial y}\frac{\partial \overline{\upsilon}}{\partial y} &= \left(\frac{\partial u}{\partial \xi}\frac{\partial \xi}{\partial y} + \frac{\partial u}{\partial \eta}\frac{\partial \eta}{\partial y}\right)\left(\frac{\partial \overline{\upsilon}}{\partial \xi}\frac{\partial \xi}{\partial y} + \frac{\partial \overline{\upsilon}}{\partial \eta}\frac{\partial \eta}{\partial y}\right) \\
&= \frac{\partial u}{\partial \xi}\frac{\partial \overline{\upsilon}}{\partial \xi}\left(\frac{\partial \xi}{\partial y}\frac{\partial \xi}{\partial y}\right) + \frac{\partial u}{\partial \eta}\frac{\partial \overline{\upsilon}}{\partial \eta}\left(\frac{\partial \eta}{\partial y}\frac{\partial \eta}{\partial y}\right) + \frac{\partial u}{\partial \xi}\frac{\partial \overline{\upsilon}}{\partial \eta}\left(\frac{\partial \xi}{\partial y}\frac{\partial \eta}{\partial y}\right) + \frac{\partial u}{\partial \eta}\frac{\partial \overline{\upsilon}}{\partial \xi}\left(\frac{\partial \xi}{\partial y}\frac{\partial \eta}{\partial y}\right) \\
&= \frac{\partial u}{\partial \xi}\frac{\partial \overline{\upsilon}}{\partial \xi}\mathcal{G}^{\xi\xi}_{yy} + \frac{\partial u}{\partial \eta}\frac{\partial \overline{\upsilon}}{\partial \eta}\mathcal{G}^{\eta\eta}_{yy} + \frac{\partial u}{\partial \xi}\frac{\partial \overline{\upsilon}}{\partial \eta}\mathcal{G}^{\xi\eta}_{yy} + \frac{\partial u}{\partial \eta}\frac{\partial \overline{\upsilon}}{\partial \xi}\mathcal{G}^{\xi\eta}_{yy}, \qquad (3.3.8)
\end{aligned}$$

introducing the short notations, $\mathcal{G}^{\xi\xi}_{xx}$, $\mathcal{G}^{\eta\eta}_{xx}$, $\mathcal{G}^{\xi\eta}_{xx}$, $\mathcal{G}^{\xi\xi}_{yy}$, $\mathcal{G}^{\eta\eta}_{yy}$, and $\mathcal{G}^{\xi\eta}_{yy}$, for the geometric factors. Using the expansion (3.3.1) for $u, \upsilon \in V_N$, we derive the discrete operator for (3.3.6) including (3.3.7)–(3.3.8) as

$$\begin{aligned}
&\mathcal{A}^N(u, \upsilon) \\
= &\sum_{e=1}^{E}\sum_{\hat{i},\hat{j}=0}^{N}\sum_{i,j=0}^{N} \upsilon^e_{ij}\left(\int_I \frac{\partial \psi_{ij}}{\partial \xi}\frac{\partial \psi_{\hat{i}\hat{j}}}{\partial \xi}\bar{\mathcal{G}}^{11}Jd\mathbf{r} + \frac{\partial \psi_{ij}}{\partial \xi}\frac{\partial \psi_{\hat{i}\hat{j}}}{\partial \eta}\bar{\mathcal{G}}^{12}Jd\mathbf{r}\right)u^e_{ij} \qquad (3.3.9) \\
&+\sum_{e=1}^{E}\sum_{\hat{i},\hat{j}=0}^{N}\sum_{i,j=0}^{N} \upsilon^e_{ij}\left(\int_I \frac{\partial \psi_{ij}}{\partial \eta}\frac{\partial \psi_{\hat{i}\hat{j}}}{\partial \xi}\bar{\mathcal{G}}^{21}Jd\mathbf{r} + \frac{\partial \psi_{ij}}{\partial \eta}\frac{\partial \psi_{\hat{i}\hat{j}}}{\partial \eta}\bar{\mathcal{G}}^{22}Jd\mathbf{r}\right)u^e_{ij}, \qquad (3.3.10)
\end{aligned}$$

where $d\mathbf{r} = d\xi d\eta$, J is the Jacobian for the coordinate transformation from the relation

$$J = \begin{vmatrix} \frac{\partial x}{\partial \xi} & \frac{\partial x}{\partial \eta} \\ \frac{\partial y}{\partial \xi} & \frac{\partial y}{\partial \eta} \end{vmatrix} \quad \text{from} \quad \begin{pmatrix} \frac{\partial x}{\partial \xi} & \frac{\partial x}{\partial \eta} \\ \frac{\partial y}{\partial \xi} & \frac{\partial y}{\partial \eta} \end{pmatrix}\begin{pmatrix} \frac{\partial \xi}{\partial x} & \frac{\partial \xi}{\partial y} \\ \frac{\partial \eta}{\partial x} & \frac{\partial \eta}{\partial y} \end{pmatrix} \equiv \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad (3.3.11)$$

and the geometric factors on each local element are defined by

$$\bar{\mathcal{G}}^{11} = (\mathcal{G}^{\xi\xi}_{xx} + \mathcal{G}^{\xi\xi}_{yy}), \qquad \bar{\mathcal{G}}^{12} = (\mathcal{G}^{\xi\eta}_{xx} + \mathcal{G}^{\xi\eta}_{yy}), \qquad (3.3.12)$$

$$\bar{\mathcal{G}}^{21} = (\mathcal{G}^{\xi\eta}_{xx} + \mathcal{G}^{\xi\eta}_{yy}), \qquad \bar{\mathcal{G}}^{22} = (\mathcal{G}^{\eta\eta}_{xx} + \mathcal{G}^{\eta\eta}_{yy}). \qquad (3.3.13)$$

Apply the numerical quadrature on the GLL nodes for the integrations in Eqs. (3.3.9) – (3.3.10) as

$$\int_I \frac{\partial \psi_{ij}}{\partial \xi}\frac{\partial \psi_{\hat{i}\hat{j}}}{\partial \xi}\bar{\mathcal{G}}^{11}d\mathbf{r} = \sum_{k,m=0}^{N} \bar{\mathcal{G}}^{11}_{km} J_{km} w_k w_m l_i'(\xi_k) l_j(\eta_m) l_{\hat{i}}'(\xi_k) l_{\hat{j}}(\eta_m), \quad (3.3.14)$$

$$\int_I \frac{\partial \psi_{ij}}{\partial \xi}\frac{\partial \psi_{\hat{i}\hat{j}}}{\partial \eta}\bar{\mathcal{G}}^{12}d\mathbf{r} = \sum_{k,m=0}^{N} \bar{\mathcal{G}}^{12}_{km} J_{km} w_k w_m l_i'(\xi_k) l_j(\eta_m) l_{\hat{i}}(\xi_k) l_{\hat{j}}'(\eta_m), \quad (3.3.15)$$

$$\int_I \frac{\partial \psi_{ij}}{\partial \eta}\frac{\partial \psi_{\hat{i}\hat{j}}}{\partial \xi}\bar{\mathcal{G}}^{21}d\mathbf{r} = \sum_{k,m=0}^{N} \bar{\mathcal{G}}^{21}_{km} J_{km} w_k w_m l_i(\xi_k) l_j'(\eta_m) l_{\hat{i}}'(\xi_k) l_{\hat{j}}(\eta_m), \quad (3.3.16)$$

$$\int_I \frac{\partial \psi_{ij}}{\partial \eta}\frac{\partial \psi_{\hat{i}\hat{j}}}{\partial \eta}\bar{\mathcal{G}}^{22}d\mathbf{r} = \sum_{k,m=0}^{N} \bar{\mathcal{G}}^{22}_{km} J_{km} w_k w_m l_i(\xi_k) l_j'(\eta_m) l_{\hat{i}}(\xi_k) l_{\hat{j}}'(\eta_m), \quad (3.3.17)$$

where $\bar{\mathcal{G}}^{(\cdot)}_{km}$ and J_{km} represent the geometric values and the Jacobian at the nodal points, respectively, and w_k and w_m are the one-dimensional GLL quadrature weights. Note that $\bar{\mathcal{G}}^{12}_{km} = \bar{\mathcal{G}}^{21}_{km}$. Finally, we have (3.3.6) in a discrete form as the following:

$$\mathcal{A}^N(u,v) = \sum_{e=1}^{E} (v^e)^T \begin{bmatrix} \mathbf{D}_x \\ \mathbf{D}_y \end{bmatrix}^T \begin{bmatrix} \mathbf{G}^{11} & \mathbf{G}^{12} \\ \mathbf{G}^{21} & \mathbf{G}^{22} \end{bmatrix}^e \begin{bmatrix} \mathbf{D}_x \\ \mathbf{D}_y \end{bmatrix} u^e \quad (3.3.18)$$

$$= \sum_{e=1}^{E} (v^e)^T \mathbf{D}^T \mathbf{G}^e \mathbf{D} u^e, \quad (3.3.19)$$

where the differentiation matrices with respect to ξ and η, $\mathbf{D}_\xi$ and $\mathbf{D}_\eta$, respectively, are written as

$$\mathbf{D}_\xi = \mathbf{I} \otimes \hat{\mathbf{D}} \text{ and } \mathbf{D}_\eta = \hat{\mathbf{D}} \otimes \mathbf{I} \quad (3.3.20)$$

in a tensor product form of the one-dimensional differentiation matrix $\hat{\mathbf{D}}_{ki} := l_i'(\xi_k)$ for $i,k = 0,1,...,N$ with the identity matirx $\mathbf{I} \in R^{(N+1)\times(N+1)}$. The differentiation matrix elements are

$$\hat{\mathbf{D}}_{ij} = \frac{L_N(\xi_i)}{L_N(\xi_j)(\xi_i - \xi_j)} \ (i \neq j); \ \hat{\mathbf{D}}_{00} = -\frac{(N+1)N}{4}; \ \hat{\mathbf{D}}_{NN} = \frac{(N+1)N}{4}; \ \hat{\mathbf{D}}_{ii} = 0,$$

which is skew-centrosymmetric $\hat{\mathbf{D}}_{ij} = -\hat{\mathbf{D}}_{N-i,N-j}$. Note that (3.3.18) involves the pointwise multiplication of the nodal values $u^e = [u^e_l]$ by each diagonal component of

$\mathbf{G}^{(\cdot)} = [\mathbf{G}_l^{(\cdot)}] = \text{diag}\{\bar{\mathcal{G}}_{km}^{(\cdot)} J_{km} w_k w_m\}$ for $l = k + (N+1)(m-1)$ on the nodal points on a local element Ω^e. Let us denote the stiffness matrix on Ω as $\mathbf{A}$, using the local stiffness matrices $\mathbf{A}^e$, represented by

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}^1 & & & \\ & \ddots & & \\ & & \mathbf{A}^e & \\ & & & \ddots \\ & & & & \mathbf{A}^E \end{bmatrix} \quad \text{with} \quad \mathbf{A}^e = \mathbf{D}^T \mathbf{G}^e \mathbf{D}, \tag{3.3.21}$$

so that we can simply write Eq. (3.3.18) as

$$\mathcal{A}^N(u, v) = \mathbf{v}^T \mathbf{A} \mathbf{u}. \tag{3.3.22}$$

Arithmetic Operations: The matrix $\mathbf{A}$ is never explicitly formed. We perform matrix-matrix multipication only acting on the block diagonal matrices $\mathbf{A}^e$. Let u_{ij}^e be arranged in column-wise consecutive entries, denoted as `u(i,j,e)`, or simply $\mathtt{u}^e$. We begin with the tensor product based derivative evaluations (3.3.18) that can be recast as matrix-matrix products for each element e as the following:

$$\mathtt{u}_\xi := (\mathbf{I} \otimes \hat{\mathbf{D}}) u^e := \hat{\mathbf{D}}_{ik} \mathtt{u}(\mathtt{k}, \mathtt{j}, \mathtt{e}) := \hat{\mathbf{D}} \mathtt{u}^e, \tag{3.3.23}$$

$$\mathtt{u}_\eta := (\hat{\mathbf{D}} \otimes \mathbf{I}) u^e := \mathtt{u}(\mathtt{i}, \mathtt{k}, \mathtt{e}) \hat{\mathbf{D}}_{kj}^T := \mathtt{u}^e \hat{\mathbf{D}}^T, \tag{3.3.24}$$

where

$$\hat{\mathbf{D}} = \begin{bmatrix} \hat{\mathbf{D}}_{00} & \hat{\mathbf{D}}_{01} & ... & \hat{\mathbf{D}}_{0N} \\ \vdots & \vdots & \vdots & \vdots \\ \hat{\mathbf{D}}_{N0} & \hat{\mathbf{D}}_{N1} & ... & \hat{\mathbf{D}}_{NN} \end{bmatrix} \quad \text{and} \quad \mathtt{u}^e = \begin{bmatrix} u_{00}^e & u_{01}^e & ... & u_{0N}^e \\ \vdots & \vdots & \vdots & \vdots \\ u_{N0}^e & u_{N1}^e & ... & u_{NN}^e \end{bmatrix}, \tag{3.3.25}$$

which requires $2E(N+1)^3$ operations on Ω. The pointwise multiplications with the geometric factors $\mathtt{u}_x = \mathbf{G}^{11}\mathtt{u}_\xi + \mathbf{G}^{12}\mathtt{u}_\eta$ and $\mathtt{u}_y = \mathbf{G}^{21}\mathtt{u}_\xi + \mathbf{G}^{22}\mathtt{u}_\eta$ require $6E(N+1)^2$ operations. Then we compute the summation of transposed derivative operators, $\hat{\mathbf{D}}^T\mathtt{u}_x + \hat{\mathbf{D}}^T\mathtt{u}_y$, involving $4E(N+1)^3 + E(N+1)^2$ operations. Thus the total operation count is $6E(N+1)^3 + 7E(N+1)^2$. The leading-order storage requirement for the factored stiffness matrix is $3E(N+1)^2$ due to the relation $\mathbf{G}_{12} = \mathbf{G}_{21}$ on Ω^e.

Direct Stiffness Summation: The solution vector in (3.3.22) is based on the unassembled representation, recalling Figure 3.2(a), without applying element interface continuity. To construct the solution vector to be continuous across element interfaces on the coincident nodal values in the assembled representation, as shown in Figure 3.2(b):

$$(x_{ij}, y_{ij})^e = (x_{\hat{i}\hat{j}}, y_{\hat{i}\hat{j}})^{\hat{e}} \rightarrow u^e_{ij} = u^{\hat{e}}_{\hat{i}\hat{j}} \;\text{ for }\; e \neq \hat{e}, \tag{3.3.26}$$

we introduce a Boolean connectivity matrix $\mathbf{Q}$ [47] that maps the local representation $\mathbf{u}$ to the global $\underline{\mathbf{u}}$, referred as the scatter operation, whereas $\mathbf{Q}^T$ acts as the gather operation:

$$\mathbf{u} = \mathbf{Q}\underline{\mathbf{u}} \;\text{ and }\; \mathbf{u}^* = \mathbf{Q}^T\mathbf{u}. \tag{3.3.27}$$

The action of $\mathbf{Q}$ on $\underline{\mathbf{u}}$ returns the copy entries of $\underline{\mathbf{u}}$ on the coincident nodes, whereas the action of $\mathbf{Q}^T$ on $\mathbf{u}$ returns $\mathbf{u}^*$ with the sum entries of $\mathbf{u}$ on the coincident nodes. The interior nodes are unchanged from both of the actions. In the assembled representation after applying the continuity, we can express Eq. (3.3.22) as

$$\mathcal{A}^N(u, v) = \underline{\mathbf{v}}^T\mathbf{Q}^T\mathbf{A}\mathbf{Q}\underline{\mathbf{u}} = \underline{\mathbf{v}}^T\bar{\mathbf{A}}\underline{\mathbf{u}}. \tag{3.3.28}$$

In practical implementations, we write our algorithms in an element-based format, utilizing local matrix-vector products evaluated independently. We use the gather–scatter operation, referred as *direct stiffness summation*, or simply *dssum*, based on a local-to-global mapping array handling the actions of $\mathbf{Q}$ and $\mathbf{Q}^T$ without constructing $\mathbf{Q}$ and $\mathbf{Q}^T$ explicitly. The detailed description on the algorithms and parallel implementations for the gather–scatter operation can be found in Chapter 4 and Chapter 8 in [47]. Here we denote the gather–scatter operation by

$$\tilde{\sum} := \mathbf{Q}\mathbf{Q}^T. \tag{3.3.29}$$

For a continuous function u and its numerical approximation $\mathbf{u}$ in the local ordering representation, the following expressions are equivalent:

$$\mathbf{Q}^T\mathbf{A}\mathbf{Q}\underline{\mathbf{u}} \iff \tilde{\sum}\mathbf{A}\mathbf{u} := \left(\mathbf{Q}\mathbf{Q}^T\right)\mathbf{A}\mathbf{u}. \tag{3.3.30}$$

We note that the local-to-local transformation $\mathbf{QQ}^T$ can be viewed as a single operation, involving summation of the variables on the shared interface nodes and redistribution of them to their original locations within one communication. To perform (3.3.28), we first compute (3.3.22) on the local data $\mathbf{u}$ and then apply the *dssum* operation as represented in (3.3.30).

3.3.2 Mass Matrices

To obtain the mass matrix, we consider the following inner product:

$$\mathcal{B}(u,v) = \int_\Omega u\overline{v}d\Omega, \tag{3.3.31}$$

which can be discretized as

$$\begin{aligned}
\mathcal{B}^N(u,v) &= \sum_{e=1}^{E}\sum_{\hat{i},\hat{j}=0}^{N}\sum_{i,j=0}^{N} v_{ij}^e\left(\int_{\Omega^e}\psi_{ij}\psi_{\hat{i}\hat{j}}d\Omega\right)u_{ij}^e \\
&= \sum_{e=1}^{E}\sum_{\hat{i},\hat{j}=0}^{N}\sum_{i,j=0}^{N} v_{ij}^e\left(\int_{I}\psi_{ij}\psi_{\hat{i}\hat{j}}Jd\mathbf{r}\right)u_{ij}^e \\
&= \sum_{e=1}^{E}\sum_{\hat{i},\hat{j}=0}^{N}\sum_{i,j=0}^{N} v_{ij}^e\left(\sum_{k,m=0}^{N} J_{km}w_kw_ml_i(\xi_k)l_j(\eta_m)l_{\hat{i}}(\xi_k)l_{\hat{j}}(\eta_m)\right)u_{ij}^e \\
&= \sum_{e=1}^{E}(v^e)^T\mathbf{J}^e\left(\hat{\mathbf{M}}\otimes\hat{\mathbf{M}}\right)u^e,
\end{aligned} \tag{3.3.32}$$

where $\hat{\mathbf{M}} = \mathrm{diag}\{w_k\}$ is the one-dimensional mass matrix and $\mathbf{J}^e = [\mathbf{J}_{ll}^e] = \mathrm{diag}\{J_{km}\}$ for $l = k + (N+1)(m-1)$. We can denote the mass matrix $\mathbf{B}$ on Ω with the local mass matrices $\mathbf{B}^e \in R^{(N+1)(N+1)}$, defined by

$$\mathbf{B} = \begin{bmatrix} \mathbf{B}^1 & & & \\ & \ddots & & \\ & & \mathbf{B}^e & \\ & & & \ddots \\ & & & & \mathbf{B}^E \end{bmatrix} \quad \text{with} \quad \mathbf{B}^e = \mathbf{J}^e(\hat{\mathbf{M}}\otimes\hat{\mathbf{M}}), \tag{3.3.33}$$

so that we can simply write Eq. (3.3.32) as

$$\mathcal{B}^N(u,v) = \mathbf{v}^T\mathbf{B}\mathbf{u}. \tag{3.3.34}$$

The Eq. (3.3.34) in the assembled representation after applying the continuity can be expressed as

$$\mathcal{B}^N(u,v) = \underline{\mathbf{v}}^T\mathbf{Q}^T\mathbf{B}\mathbf{Q}\underline{\mathbf{u}} = \underline{\mathbf{v}}^T\bar{\mathbf{B}}\underline{\mathbf{u}}. \tag{3.3.35}$$

3.3.3 Quasi-Periodic Matrix

Consider the following inner product for the quasi-periodic operator in Eq. (3.2.33):

$$C(u,v) = \int_\Omega \frac{\partial u}{\partial x}\overline{v}d\Omega, \tag{3.3.36}$$

which can be discretized as

$$\begin{aligned}
\mathcal{C}^N(u,v) &= \sum_{e=1}^{E}\sum_{\hat{i},\hat{j}=0}^{N}\sum_{i,j=0}^{N} v_{ij}^e\left(\int_{\Omega^e}\frac{\partial\psi_{ij}}{\partial x}\psi_{\hat{i}\hat{j}}d\Omega\right)u_{ij}^e \\
&= \sum_{e=1}^{E}\sum_{\hat{i},\hat{j}=0}^{N}\sum_{i,j=0}^{N} v_{ij}^e\left(\int_{I}\frac{\partial\psi_{ij}}{\partial x}\psi_{\hat{i}\hat{j}}Jd\mathbf{r}\right)u_{ij}^e \\
&= \sum_{e=1}^{E}\sum_{\hat{i},\hat{j}=0}^{N}\sum_{i,j=0}^{N} v_{ij}^e\left(\sum_{k,m=0}^{N} J_{km}w_kw_m l_i'(\xi_k)l_j(\eta_m)l_{\hat{i}}(\xi_k)l_{\hat{j}}(\eta_m)\right)u_{ij}^e \\
&= \sum_{e=1}^{E}(v^e)^T\mathbf{J}^e\left(\hat{\mathbf{M}}\otimes\hat{\mathbf{D}}\right)u^e.
\end{aligned} \tag{3.3.37}$$

In convention, (3.3.36) is referred as the convective operator. In this context, we paticularly refer it as the quasi-periodic operator, due to that the operator is a derivative, resulting from imposing the periodicity into the solution as in (3.2.30) for the oblique incident ($\alpha \neq 0$). We can express the quasi-periodic matrix on Ω using the local quasi-periodic matrices $\mathbf{C}^e \in R^{(N+1)(N+1)}$:

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}^1 & & & \\ & \ddots & & \\ & & \mathbf{C}^e & \\ & & & \ddots \\ & & & & \mathbf{C}^E \end{bmatrix} \quad \text{with} \quad \mathbf{C}^e = \mathbf{J}^e(\hat{\mathbf{M}}\otimes\hat{\mathbf{D}}), \tag{3.3.38}$$

and we can simply write Eq. (3.3.37) as

$$\mathcal{C}^N(u,\upsilon) = \mathbf{v}^T\mathbf{C}\mathbf{u}. \tag{3.3.39}$$

The Eq. (3.3.39) in the assembled representation after applying the continuity can be expressed as

$$\mathcal{C}^N(u,\upsilon) = \underline{\mathbf{v}}^T\mathbf{Q}^T\mathbf{C}\mathbf{Q}\underline{\mathbf{u}} = \underline{\mathbf{v}}^T\bar{\mathbf{C}}\underline{\mathbf{u}}. \tag{3.3.40}$$

3.3.4 Dirichlet-to-Neumann (DtN) Boundary Discretization

In this chapter, our formulation is based on the case with DtN boundaries in y. Similar approach can be applied for the other case with DtN boundaries in x, by simply changing the variable. Let us define our domputational domain as $\Omega = [0, L] \times [a, b]$ with L–periodicity in x and DtN boundaries in y. Consider the DtN boundary surface $\Gamma = \{(x, b) \in \partial\Omega \mid x \in [0, L]\}$ where $\partial\Omega$ represents the boundary of Ω. Let us denote $\Gamma = \cup_{\hat{e}=1}^{\hat{E}}\Gamma^{\hat{e}}$ where $\Gamma^{\hat{e}} = \Omega^{\hat{e}} \cap \partial\Omega$ are nonoverlapping DtN boundary surfaces on the local elements $\Omega^{\hat{e}}$ containing the DtN boundary surfaces. We define a dtn-to-local mapping array that contain the indices of the DtN surface nodes $(i, j, \hat{e})$ to the local index $(i, j, e) :=$ dtn-to-local$(i, j, \hat{e})$. We note that DtN boundary nodes in y fall on the index either with $j = 0$ or with $j = N$, which will be simply represented by a fixed index as $j = j_b$.

DtN Matrix T: We can represent our approxiamte solution on $\Gamma^{\hat{e}}$ in the form of (3.3.1) as

$$u^{\hat{e}}(x, b) = \sum_{i,j=0}^{N} u_{ij}^{\hat{e}} l_i(\xi) l_j(\eta(b)) = \sum_{i=0}^{N} u_{ij_b}^{\hat{e}} l_i(\xi). \tag{3.3.41}$$

For the surface integration in Eq. (3.2.34) with Eqs. (3.2.23)–(3.2.24), we have

$$\begin{aligned} T(u,\upsilon)_\Gamma &= \int_\Gamma T[u]\overline{\upsilon}d\Gamma \\ &= \int_\Gamma \left(\sum_{p=-\infty}^{\infty} \mathbf{i}\beta_p \hat{u}_p e^{\mathbf{i}d_p x}\right)\overline{\upsilon}dx \\ &= \sum_{p=-\infty}^{\infty} \mathbf{i}\beta_p \hat{u}_p \int_\Gamma e^{\mathbf{i}d_p x}\overline{\upsilon}dx, \end{aligned} \tag{3.3.42}$$

where $d_p = \frac{2\pi p}{d}$ and $\hat{u}_p$ are the one-dimensional Fourier coefficients of $u(x,b)$ on Γ given as

$$\hat{u}_p = \frac{1}{L}\int_0^L u(x',b)e^{-\mathbf{i}d_p x'}dx' \approx \frac{1}{L}\sum_{\hat{e}=1}^{\hat{E}}\int_{\Gamma^{\hat{e}}} u^{\hat{e}}(x',b)e^{-\mathbf{i}d_p x'}dx'. \tag{3.3.43}$$

Plugging (3.3.43) into (3.3.42) with a finite expansion of $T[u]$ ($|p| \le P$) and applying (3.3.41), we have

$$\begin{aligned} & T^N(u,\upsilon)_\Gamma \\ = & \sum_{p=-P}^{P} \mathbf{i}\beta_p \left(\frac{1}{L}\sum_{\hat{e}=1}^{\hat{E}}\int_{\Gamma^{\hat{e}}} u^{\hat{e}}(x',b)e^{-\mathbf{i}d_p x'}dx'\right)\left(\sum_{\bar{e}=1}^{\hat{E}}\int_{\Gamma^{\bar{e}}} e^{\mathbf{i}d_p x}\overline{\upsilon}dx\right) & (3.3.44) \\ = & \sum_{p=-P}^{P} \mathbf{i}\beta_p \left[\sum_{i=0}^{N} u_{ij_b}^{\hat{e}}\left(\frac{1}{L}\sum_{\hat{e}=1}^{\hat{E}}\int_{\Gamma^{\hat{e}}} l_i(\xi)e^{-\mathbf{i}d_p x'}dx'\right)\right]\left(\sum_{\bar{e}=1}^{\hat{E}}\int_{\Gamma^{\bar{e}}} e^{\mathbf{i}d_p x}\overline{\upsilon}dx\right) & (3.3.45) \end{aligned}$$

Choosing $\overline{\upsilon} = l_{\hat{i}}(\xi)$ with a different index set of $\hat{i}$ on each $\Omega^{\hat{e}}$ and defining the following:

$$s_i^{\hat{e},p} = \frac{1}{\sqrt{L}}\int_{\Gamma^{\hat{e}}} l_i(\xi)e^{-\mathbf{i}d_p x'}\,\mathrm{d}x' \quad \text{and} \quad s_{\hat{i}}^{\bar{e},-p} = \frac{1}{\sqrt{L}}\int_{\Gamma^{\bar{e}}} l_{\hat{i}}(\xi)e^{\mathbf{i}d_p x}\,\mathrm{d}x, \tag{3.3.46}$$

we can express (3.3.42) in a simplified form as

$$T^N(u,\upsilon)_\Gamma = \sum_{\hat{e}=1}^{\hat{E}}\sum_{i=0}^{N} u_{ij_b}^{\hat{e}}\left[\sum_{p=-P}^{P}\mathbf{i}\beta_p\left(s_i^{\hat{e},p}\right)\left(\sum_{\bar{e}=1}^{\hat{E}} s_{\hat{i}}^{\bar{e},-p}\right)\right] = \sum_{\hat{e}=1}^{\hat{E}}\sum_{i=0}^{N} u_{ij_b}^{\hat{e}} T_{\hat{i}i}^{\hat{e}}. \tag{3.3.47}$$

Here we note that $s_i^{\hat{e},-p}$ is the complex conjugate of $s_i^{\hat{e},p}$ from the following:

$$\overline{s_i^{\hat{e},p}} = \overline{\frac{1}{\sqrt{L}}\int_{\Gamma^{\hat{e}}} l_i(\xi)e^{-\mathbf{i}d_p x}\,\mathrm{d}x} = \frac{1}{\sqrt{L}}\int_{\Gamma^{\hat{e}}} l_i(\xi)e^{\mathbf{i}d_p x}\,\mathrm{d}x = s_i^{\hat{e},-p}. \tag{3.3.48}$$

Thus we only need to compute $s_i^{\hat{e},p}$ for $p \geq 0$ to obtain:

$$T_{\hat{i}i}^{\hat{e}} = \mathbf{i}\left(\beta_0 s_i^{\hat{e},0}\sum_{\bar{e}=1}^{\hat{E}} s_{\hat{i}}^{\bar{e},0} + \sum_{p=1}^{P}(\beta_p+\beta_{-p})\left(s_i^{\hat{e},p}+\overline{s_i^{\hat{e},p}}\right)\left[\sum_{\bar{e}=1}^{\hat{E}}\overline{s_{\hat{i}}^{\bar{e},p}}\right]\right), \tag{3.3.49}$$

where $\beta_p = \beta_{-p}$ if $\alpha = 0$; $\beta_p \neq \beta_{-p}$ if $\alpha \neq 0$ and no particular relation can be found between β_p and β_{-p} in general. Here $T_{\hat{i}i}^{\hat{e}}$ is a complex number so that we can alternatively write (3.3.47) as

$$T^N(u,v)_\Gamma = \sum_{\hat{e}=1}^{\hat{E}}\sum_{i=0}^{N} u_{ij_b}^{\hat{e}} T_{\hat{i}i}^{\hat{e}} = \sum_{\hat{e}=1}^{\hat{E}}\sum_{i=0}^{N} u_{ij_b}^{\hat{e}}\left[(T_{\hat{i}i}^{\hat{e}})_{\text{real}} + \mathbf{i}(T_{\hat{i}i}^{\hat{e}})_{\text{imag}}\right]. \tag{3.3.50}$$

Now, we can map the values of $T_{\hat{i}i}^{\hat{e}}$ into a matrix $\mathbf{T}^e = \left[\mathbf{T}_{\hat{l}l}^e\right]$ for $\hat{l} = \hat{i} + (N+1)j$ and $l = i + (N+1)j$ from the dtn-to-local mapping $(\hat{i}, j, e) :=$ dtn-to-local$(\hat{i}, j_b, \hat{e})$ and $(i, j, e) :=$ dtn-to-local$(i, j_b, \hat{e})$. Similarly, $\{u_{ij_b}^{\hat{e}}\}$ can be mapped to the local data $\{u_{ij}^e\}$. Note that the entries of $\mathbf{T}^e$ are zeros if the indices are not indicating the DtN boundary nodes. Finally, we have Eq. (3.3.47) in the local representation form as

$$\mathcal{T}^N(u,v) = \sum_{e=1}^{E}(v^e)^T\mathbf{T}^e u^e = \mathbf{v}^T\mathbf{T}\mathbf{u} = \mathbf{v}^T(\mathbf{T}_{\text{r}} + \mathbf{i}\mathbf{T}_{\text{i}})\mathbf{u}, \tag{3.3.51}$$

where $\mathbf{T}_{\text{r}}$ and $\mathbf{T}_{\text{i}}$ represent the real and imaginary part of the complex matrix $\mathbf{T}$. Then we have the assembled representation of (3.3.51) as

$$\mathcal{T}^N(u,v) = \underline{\mathbf{v}}^T\mathbf{Q}^T\mathbf{T}\mathbf{Q}\underline{\mathbf{u}} = \underline{\mathbf{v}}^T\bar{\mathbf{T}}\underline{\mathbf{u}} = \underline{\mathbf{v}}^T(\bar{\mathbf{T}}_{\text{r}} + \mathbf{i}\bar{\mathbf{T}}_{\text{i}})\underline{\mathbf{u}}. \tag{3.3.52}$$

Compute Matrix T: Now we discuss how to compute $s_i^{\hat{e},p}$ in Eq. (3.3.49). Note that these data are pre-computed only one time. One might apply the GLL quadrature for the integrations when d_p is small. However, for large d_p, the GLL quadrature is not accurate enough to capture the high frequency modes. One can consider the discrete FFT algorithm since it is the pth component of the inverse DFFT of function $l_i(\xi)$. However, $l_i(\xi)$ has only very small portion of compact support on Γ so that we can compute it directly on its local compact support using refined GLL quadrature points on each $\Gamma^{\hat{e}}$. Another approach is to use the relation to the Bessel function, which can be more efficient than the other approach.

In this chapter, we discuss the computation of $s_i^{\hat{e},p}$ based on the Bessel function representation. Having written $l_i(\xi)$ in the finite expansion of the mth-order Legendre polynomials given as

$$l_i(\xi) = \sum_{m=0}^{N} (\hat{l}_i)_m L_m(\xi), \tag{3.3.53}$$

where $(\hat{l}_i)_m$ are the Legendre expansion coefficients defined by

$$(\hat{l}_i)_m = \frac{2m+1}{2} \int_{-1}^{1} l_i(\xi) L_m(\xi) d\xi. \tag{3.3.54}$$

Then, substituting (3.3.53) in (3.3.46) and using simply the notation x, instead of x', we have

$$s_i^{\hat{e},p} = \frac{1}{\sqrt{L}} \int_{\Gamma^{\hat{e}}} l_i(\xi(x)) e^{-\mathbf{i} d_p x} dx = \frac{1}{\sqrt{L}} \sum_{m=0}^{N} (\hat{l}_i)_m \left(\int_{-1}^{1} L_m(\xi) e^{-\mathbf{i} d_p x(\xi)} J_s^{\hat{e}} d\xi \right), \tag{3.3.55}$$

where $J_s^{\hat{e}}$ is the surface Jacobian on $\Gamma^{\hat{e}}$. In fact, each $\Gamma^{\hat{e}}$ is represented by an interval $[x_{\min}^{\hat{e}}, x_{\max}^{\hat{e}}]$ with the coordinate transformation by $x(\xi) = \hat{a}_e \xi + \hat{b}_e$ with $\hat{a}_e = (x_{\max}^{\hat{e}} - x_{\min}^{\hat{e}})/2$ and $\hat{b}_e = (x_{\max}^{\hat{e}} + x_{\min}^{\hat{e}})/2$, so that $J_s^{\hat{e}} \equiv \hat{a}_e$ is constant on $\Gamma^{\hat{e}}$. Then, Eq. (3.3.55) becomes

$$s_i^{\hat{e},p} = \frac{\hat{a}_e}{\sqrt{L}} \sum_{m=0}^{N} (\hat{l}_i)_m d_m^{p,\hat{e}} \quad \text{with} \quad q_m^{p,\hat{e}} = \int_{-1}^{1} L_m(\xi) e^{-\mathbf{i} d_p (\hat{a}_e \xi + \hat{b}_e)} d\xi. \tag{3.3.56}$$

Now we need to compute the two terms, $(\hat{l}_i)_m$ and $q_m^{p,\hat{e}}$, in (3.3.56). To compute $(\hat{l}_i)_m$, one might apply the GLL quadrature for the integration term in (3.3.54) as follows:

$$(\hat{l}_i)_m = \frac{2m+1}{2} \sum_{k=0}^{N} l_i(\xi_k) L_m(\xi_k) w_k = \frac{2m+1}{2} L_m(\xi_i) w_i. \tag{3.3.57}$$

An alternative approach is to evaluate (3.3.53) on the GLL grids in $[-1, 1]$, resulting in the form:

$$\mathbf{L}\hat{\mathbf{L}} = \begin{bmatrix} L_0(\xi_0) & L_1(\xi_0) & \cdots & L_m(\xi_0) \\ \vdots & \vdots & \vdots & \vdots \\ L_0(\xi_N) & L_1(\xi_N) & \cdots & L_m(\xi_N) \end{bmatrix} \begin{bmatrix} (\hat{l}_0)_0 & \cdots (\hat{l}_i)_0 & \cdots & (\hat{l}_N)_0 \\ \vdots & \vdots & \vdots & \vdots \\ (\hat{l}_0)_N & \cdots (\hat{l}_i)_N & \cdots & (\hat{l}_N)_N \end{bmatrix} \equiv \mathbf{I}, \tag{3.3.58}$$

and compute the inverse of the matrix $\mathbf{L} = [\mathbf{L}_{ji}] = [L_i(\xi_j)]$ to obtain $\hat{\mathbf{L}} = [\hat{\mathbf{L}}_{mi}] = [(\hat{l}_i)_m] = \mathbf{L}^{-1}$. To compute $q_m^{p,\hat{e}}$, we recall that the Legendre Polynomials are related to the Bessel functions as

$$\int_{-1}^{1} L_m(\xi) e^{-\mathbf{i}x\xi} d\xi = \frac{1}{\mathbf{i}^m}\sqrt{\frac{2\pi}{x}} J_{m+1/2}(x) = \frac{2}{\mathbf{i}^m} j_m(x) \quad \text{for} \quad x \in R, \tag{3.3.59}$$

where j_m is the spherical Bessel function and J_m is the ordinary Bessel function with the relation:

$$j_m(x) = \sqrt{\frac{\pi}{2x}} J_{m+1/2}(x). \tag{3.3.60}$$

Then, we can write

$$q_m^{p,\hat{e}} = \int_{-1}^{1} L_m(\xi) e^{-\mathbf{i}d_p(\hat{a}_e\xi + \hat{b}_e)} d\xi = e^{-\mathbf{i}d_p\hat{b}_e}\left(\frac{2}{\mathbf{i}^m} j_m(d_p\hat{a}_e)\right). \tag{3.3.61}$$

From (3.3.57) and (3.3.61), we have the final form of $s_i^{\hat{e},p}$ by

$$s_i^{\hat{e},p} = \frac{\hat{a}_e e^{-\mathbf{i}d_p\hat{b}_e}}{\sqrt{L}} \sum_{m=0}^{N} (\hat{l}_i)_m \left(\frac{2}{\mathbf{i}^m} j_m(d_p\hat{a}_e)\right). \tag{3.3.62}$$

3.3.5 Matrix Structures and Eigenvalues

In this section, we discuss a complete set of our SEM scheme with DtN boundary condition, provided with the matrix structures and analysis for the operators. We solve our solution in the form of a single real vector with the length of $2n$ expressed by $\underline{\mathbf{u}}^N = [u_{\mathrm{r}}^N, u_{\mathrm{i}}^N]^T$ where u_{r}^N and u_{i}^N represent real and imaginary part of the solution. We first set the following matrices:

$$\breve{\mathbf{H}} := \begin{bmatrix} \mathbf{A} - k^2\mathbf{B} & \mathbf{O} \\ \mathbf{O} & \mathbf{A} - k^2\mathbf{B} \end{bmatrix} \quad \text{and} \quad \breve{\mathbf{C}} := \begin{bmatrix} \mathbf{C} & \mathbf{O} \\ \mathbf{O} & \mathbf{C} \end{bmatrix}, \tag{3.3.63}$$

and the DtN boundary operator and the right-hand side for the boundary conditions in the matrix form can be expressed as

$$\breve{\mathbf{T}} := \begin{bmatrix} \mathbf{T}_{\mathrm{r}} & \mathbf{T}_{\mathrm{i}} \\ -\mathbf{T}_{\mathrm{i}} & \mathbf{T}_{\mathrm{r}} \end{bmatrix} \quad \text{and} \quad \breve{\mathbf{F}} := \begin{bmatrix} \mathbf{F} & \mathbf{O} \\ \mathbf{O} & \mathbf{F} \end{bmatrix}. \tag{3.3.64}$$

Our scheme in the assembled representation after applying the continuity and DtN boundaries can be expressed as

$$\bar{\bar{\mathbf{H}}}\underline{\mathbf{u}}^{\mathbf{N}} = \bar{\bar{\mathbf{F}}}, \tag{3.3.65}$$

where

$$\bar{\bar{\mathbf{H}}} := \begin{bmatrix} \bar{\mathbf{A}} - (k^2 - \alpha^2)\bar{\mathbf{B}} - 2\alpha\bar{\mathbf{C}} + \bar{\mathbf{T}}_{\mathrm{r}} & \bar{\mathbf{T}}_{\mathrm{i}} \\ -\bar{\mathbf{T}}_{\mathrm{i}} & \bar{\mathbf{A}} - (k^2 - \alpha^2)\bar{\mathbf{B}} + 2\alpha\bar{\mathbf{C}} + \bar{\mathbf{T}}_{\mathrm{r}} \end{bmatrix} \tag{3.3.66}$$

We demonstrate the matrix structures for the SEM operators in unassembled case (3.3.63) in Figure 3.3, the matrices for different set of the boundary conditions in Figure 3.4, provided with their eigenvalue distributions on the right panels. In Table 3.1, we demonstrate the condition numbers for the operators on the set of boundary conditions.

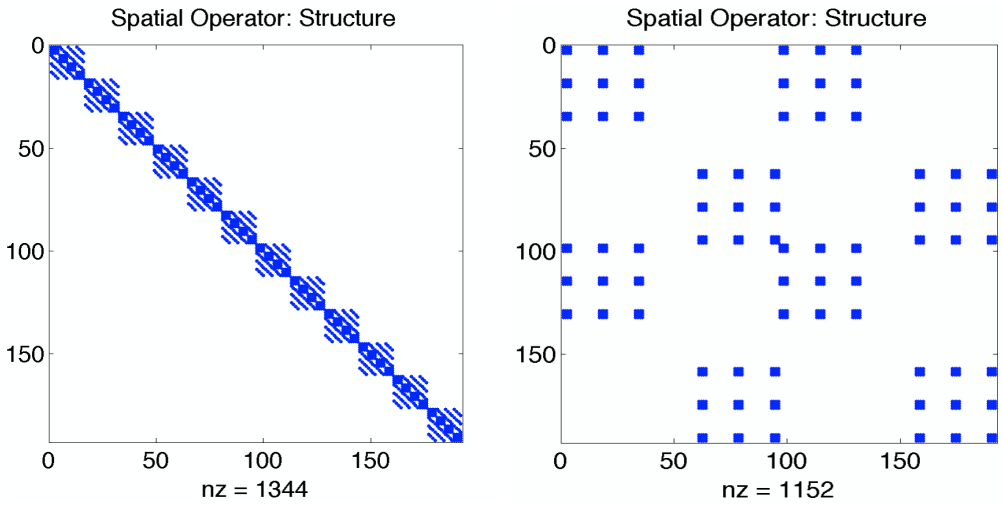
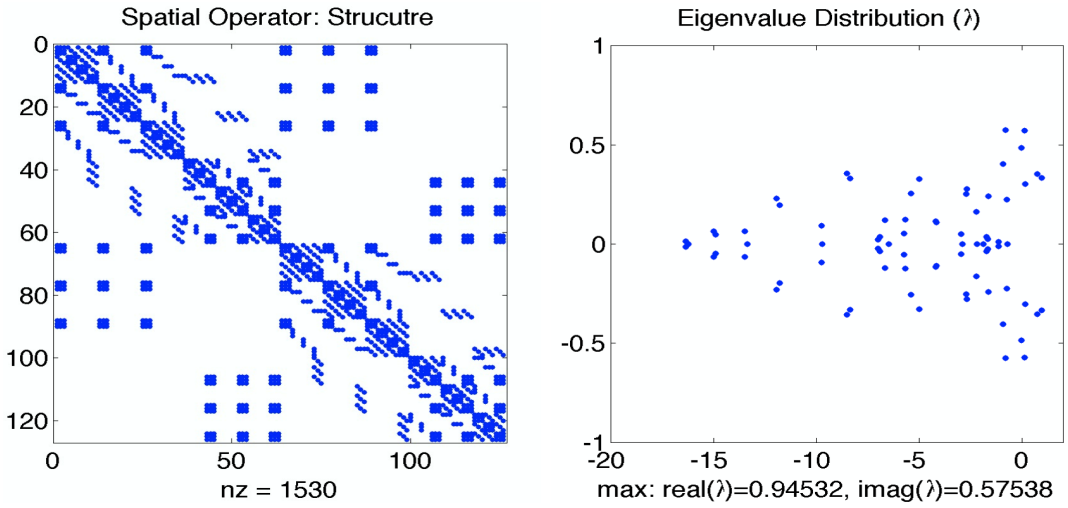


Figure 3.3. Matrix structures $\breve{\mathbf{H}}$ & $\breve{\mathbf{T}}$ (unassembled): $E = 3 \times 2$, $N = 3$.

3.3.6 Computations

In this section, we describe the Generalized Minimum RESidual (GMRES) method and provide the simple steps of our computation setup.



(a) $\bar{\bar{\mathbf{H}}}$ with DtN/DtN (top/bottom)

Figure 3.4. Matrix structures (assembled): $E = 3 \times 2$ and $N = 3$.

Table 3.1

Condition numbers

	Assembled Representations			
	DtN (top/bottom)			
	E	N	Condition #	
	3×2	3	5.954213579989568e+01	
		5	2.347021228844633e+02	
		7	5.902204731597798e+02	
		9	1.203134665072308e+03	
		11	2.143766952778180e+03	
		13	3.482577618583732e+03	

GMRES Algorithms: This algorithm approximates the solution of (3.3.73) based on the form:

$$x_m = x_0 + V_m y, \tag{3.3.67}$$

where V_m is an orthonormal basis for the Krylov subspace of dimension m defined by

$$\mathcal{K}_m = \text{span}\{r_0, Ar_0, ..., A^{m-1}r_0\}, \tag{3.3.68}$$

where $y \in R^m$ is determined to be the 2-norm of the residual $r_m = b - Ax_m$ is minimal over $\mathcal{K}_m$. Let $v_1 = r_0/\beta$ where $\beta = \|r_0\|_2$. The residual r_m associated with (3.3.73) is

$$\begin{aligned} r_m &= b - Ax_m = b - A(x_0 + V_m y) && (3.3.69)\\ &= r_0 - AV_m y = r_0 - V_{m+1}\bar{H}_m y && (3.3.70)\\ &= \beta v_1 - V_{m+1}\bar{H}_m y && (3.3.71)\\ &= V_{m+1}(\beta e_1 - \bar{H}_m y). && (3.3.72) \end{aligned}$$

Computation Algorithms: We solve the sparce non Hermitian linear systems resulting from our spectral element method in combination with the DtN boundary treatment:

$$\mathbf{H}\begin{bmatrix} \underline{u}_{\text{r}}^N \\ \underline{u}_{\text{i}}^N \end{bmatrix} = \mathbf{f}, \tag{3.3.73}$$

where $\underline{u}_{\text{r}}^N$ and $\underline{u}_{\text{i}}^N$ represent real and imaginary solutions, respectively. We denote our solution vector as $\underline{\mathbf{u}}^N = [\underline{u}_{\text{r}}^N, \underline{u}_{\text{i}}^N]^T$. In the linear system (3.3.73), our Helmholtz operator $\bar{H}$ is a square nonsingular $2n \times 2n$ real matrix and $\mathbf{f}$ is a real vector of length $2n$.

1. Set the incident field
2. Set the normal directional derivative of the incident on the DTN boundary
3. Transform $w = e^{-i\alpha d}u$.

4. Set a mask array for Dirichlet boundary conditions
5. Set boundary conditions on DtN boundaries
6. Compute DtN operator
7. Perform *dssum* for $\mathbf{f}$
8. Perform GMRES iterations
9. Perform *dssum* for $ax = \bar{\mathbf{H}}\underline{\mathbf{u}}$
10. Transform back $u = e^{i\alpha d}w$.

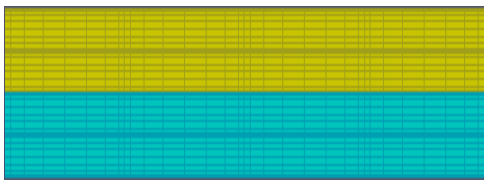
3.4 Computational Results

In this section, we demonstrate computational results for acoustic time-harmonic scattering problems in double-layer media. We consider example problems featured with geometries in three groups, smooth flat structures with exact solutions, smooth curved structures, and nonsmooth interface structures in double-layer media.

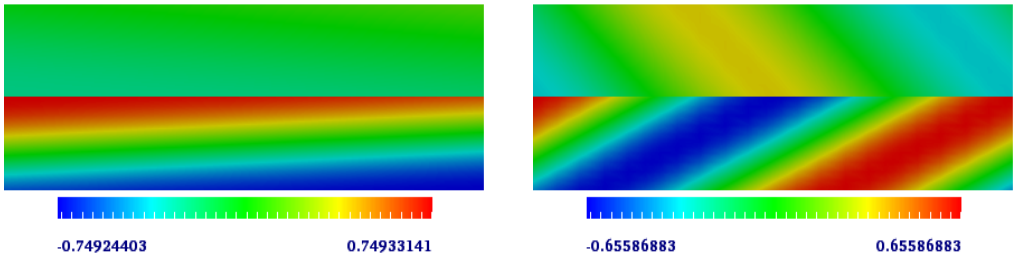
3.4.1 Smooth Flat Structures with Exact Solutions

In this section we consider DtN boundaries at the bottom and top in y. We consider scattering problems that exist analytic solutions for arbitrary incident waves and we validate our computational results, provided with convergence studies in comparison with the analytic solutions.

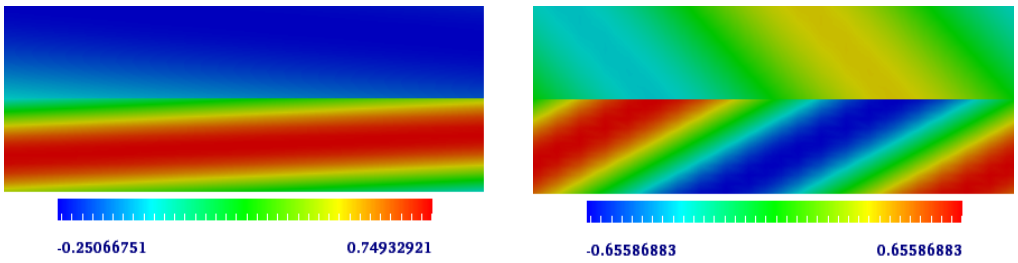
Double Layer: Consider 2π-periodicity in x for $(x, y) \in \Omega = [0, 2\pi] \times [-1, 1]$ with DtN boundaries at $y = -1$ and $y = 1$. We solve the numerical solution $\underline{\mathbf{u}}^N$ for the total field on the mesh with $E = 4 \times 4$. Figure 3.5(a) demonstrates the mesh with the GLL grids for $N = 8$. The analytic solution for the total field and the incident field are given as the following in Region 1 and Region 2:



(a) Mesh and GLL grids: $E = 4 \times 4$, $N = 8$

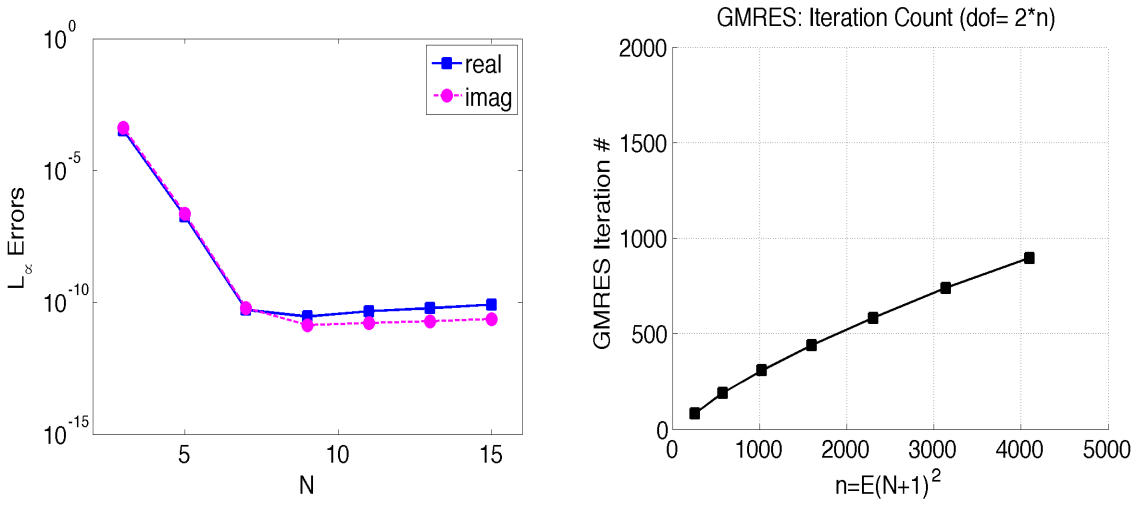


(b) Real part of the scattered field $\underline{\mathbf{u}}^N_{\text{scat}}$: $\alpha = 0.1$ (left) and $\alpha = 1.0$ (right)

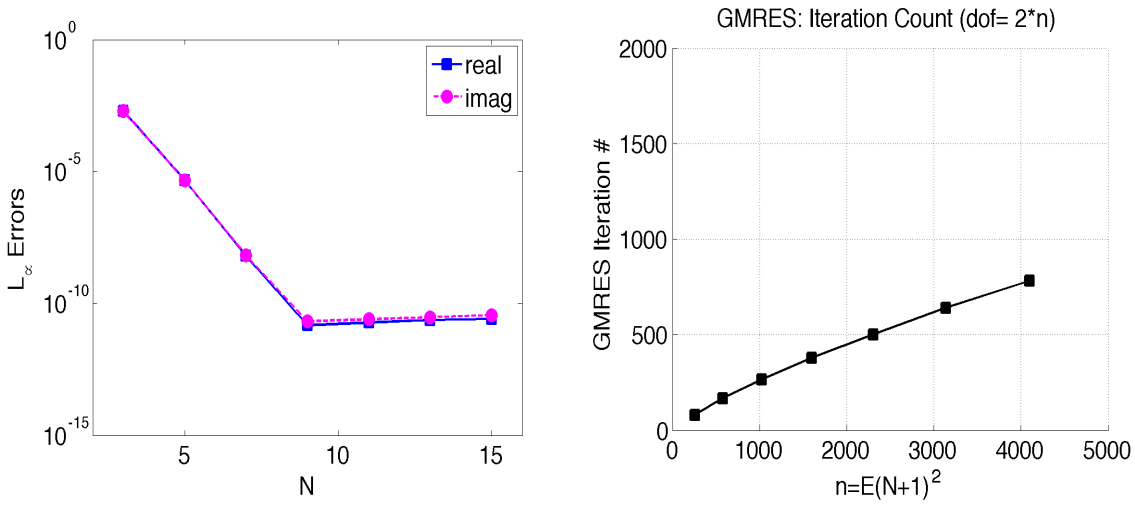


(c) Imaginary part of the scattered field $\underline{\mathbf{u}}^N_{\text{scat}}$: $\alpha = 0.1$ (left) and $\alpha = 1.0$ (right)

Figure 3.5. Impedance on flat grating: $k = 1.5$ (yellow); $k = 2.5$ (blue); DtN (top/bottom).



(a) Flat Single Layer



(b) Flat Double Layer

Figure 3.6. Convergence, GMRES iteration counts, and mesh; E=4×4 and $N = 3, 5, 7, 9, 11, 13, 15$.

- Region 1: $[0, 2\pi] \times [0, 1]$ with $\beta = \sqrt{k_2^2 - \alpha^2}$.

$$u_{\mathrm{r}}^{\mathrm{exact}}(x, y) = \cos(\alpha x - \beta y) + c_1 \cos(\alpha x + \beta y) \tag{3.4.1}$$

$$u_{\mathrm{i}}^{\mathrm{exact}}(x, y) = \sin(\alpha x - \beta y) + c_1 \sin(\alpha x + \beta y) \tag{3.4.2}$$

$$u_{\mathrm{r}}^{\mathrm{inc}}(x, y) = \cos(\alpha x - \beta y) \tag{3.4.3}$$

$$u_{\mathrm{i}}^{\mathrm{inc}}(x, y) = \sin(\alpha x - \beta y), \tag{3.4.4}$$

  where $c_1 = \frac{\hat{k}_2 - \hat{k}_1}{(\hat{k}_2 - \hat{k}_1)}$; $\hat{k}_1 = \sqrt{k_1^2 - \alpha^2}$ and $\hat{k}_2 = \sqrt{k_2^2 - \alpha^2}$ with $k_1 = 2.5$ and $k_2 = 1.5$.

- Region 2: $[0, 2\pi] \times [-1, 0]$ with $\beta = \sqrt{k_1^2 - \alpha^2}$.

$$u_{\mathrm{r}}^{\mathrm{exact}}(x, y) = d_1 \cos(\alpha x - \beta y) \tag{3.4.5}$$

$$u_{\mathrm{i}}^{\mathrm{exact}}(x, y) = d_1 \sin(\alpha x - \beta y) \tag{3.4.6}$$

  where $d_1 = \frac{2\hat{k}_2}{(\hat{k}_2 + \hat{k}_1)}$; $\hat{k}_1 = \sqrt{k_1^2 - \alpha^2}$ and $\hat{k}_2 = \sqrt{k_2^2 - \alpha^2}$ with $k_1 = 2.5$ and $k_2 = 1.5$.

We examine the cases with varying incient angle $\alpha = 0.1$ and $\alpha = 1.0$. Figures 3.5(b)–3.5(c) show our numerical solution for the scattered field. Figure 3.6 demonstrates the convergence of our numerical solutions using the maximum errors for the scattered field in the single- and double-layer media:

$$\|u_{\mathrm{scat}}^{\mathrm{exact}} - \underline{\mathbf{u}}_{\mathrm{scat}}^{N}\|_{\infty}, \tag{3.4.7}$$

where $u_{\mathrm{scat}}^{\mathrm{exact}} = u^{\mathrm{exact}} - u^{\mathrm{inc}}$ is the exact solution for the scattered field. They show spectral convergence as N increases. Due to the condition number increasing as N increases, our solution reaches to 1e-10 level at best when using the GMRES algorithm with the interation count increasing up to $\sim$ 900 for $N = 15$ as demonstrated in Figures 3.6(a)–3.6(b).

3.4.2 Smooth Curved Structures

In this section, we consider DtN boundaries at the bottom and top in y. Due to that there are no analytic solutions for these cases, we validate our results in comparison to the results by the TFE method [19].

Double Layer: Consider 2π-periodicity in x for $(x, y) \in \Omega = [0, 2\pi] \times [-1, 1]$ including two different medium: Region 1= $[0, 2\pi] \times [g(x), 1]$ and Region 2=$[0, 2\pi] \times [-1, g(x)]$ where $g(x) = \epsilon \cos(x)$ with $\epsilon = 0.1$, with DtN boundaries at $y = -1$ and $y = 1$. We solve the numerical solution $\underline{\mathbf{u}}^N$ for the total field on the mesh with $E = 4 \times 4$.

Figure 3.7(a) demonstrates the mesh with the GLL grids for $N = 8$. T The incident field are given as the following in Region 1:

$$u_{\mathrm{r}}^{\mathrm{inc}}(x, y) = \cos(\alpha x - \beta y) \tag{3.4.8}$$

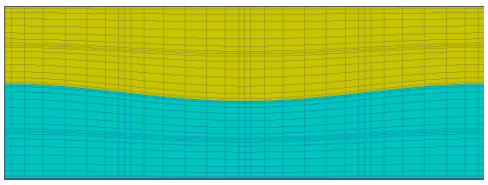
$$u_{\mathrm{i}}^{\mathrm{inc}}(x, y) = \sin(\alpha x - \beta y), \tag{3.4.9}$$

where $\beta = \sqrt{k^2 - \alpha^2}$. We examine the case of $k = k_1 = 1.5$ on $[0, 2\pi] \times [g(x), 1]$ and with $k = k_2 = 2.5$ on $[0, 2\pi] \times [-1, g(x)]$ for varying incident angle with $\alpha = 0.1$ and $\alpha = 1.0$.

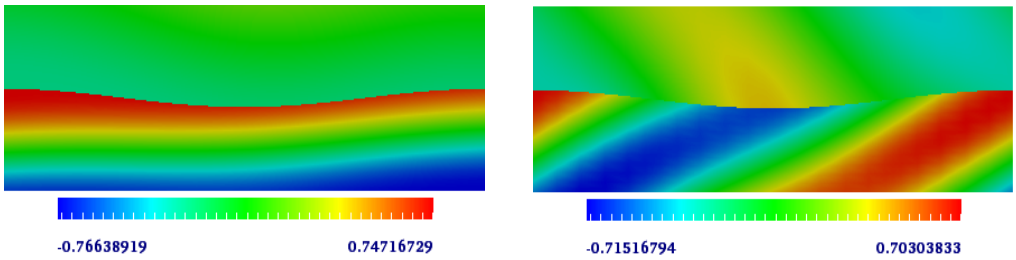
Figures 3.7(b)–3.7(c) show our numerical solution for the scattered field. Figure 3.8 demonstrates the convergence of our numerical solutions using the maximum errors for the scattered field in the single- and double-layer media in comparison with the results by the TFE:

$$\|u_{\mathrm{scat}}^{\mathrm{TFE}} - \underline{\mathbf{u}}_{\mathrm{scat}}^N\|_\infty, \tag{3.4.10}$$

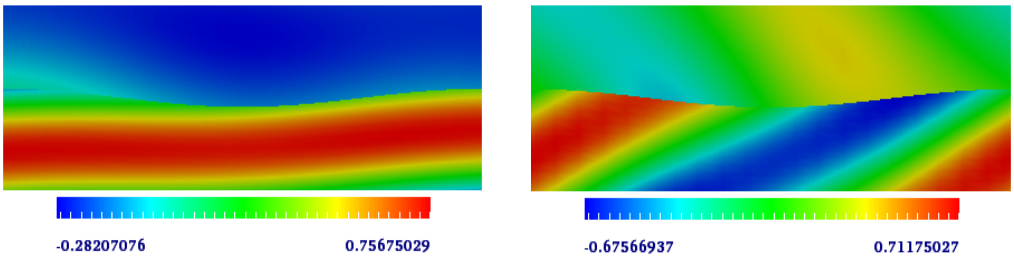
where $u_{\mathrm{scat}}^{\mathrm{TFE}} = u^{\mathrm{TFE}} - u^{\mathrm{inc}}$ is the exact solution for the scattered field. They show spectral convergence as N increases. Due to the condition number increasing as N increases, our solution reaches to 1e-10 level at best when using the GMRES algorithm with the interation count increasing up to $1700 \sim 1900$ for $N = 15$ as demonstrated in Figures 3.8(a)–3.8(b).



(a) Mesh and GLL grids: $E = 4 \times 4$, $N = 8$

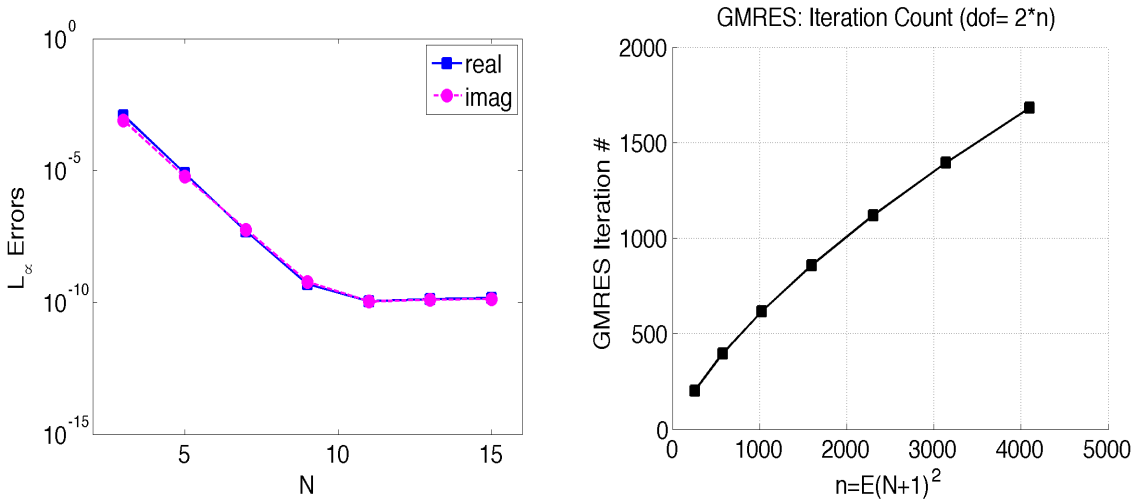


(b) Real Part: $\alpha = 0.1$ (left) and $\alpha = 1.0$ (right)

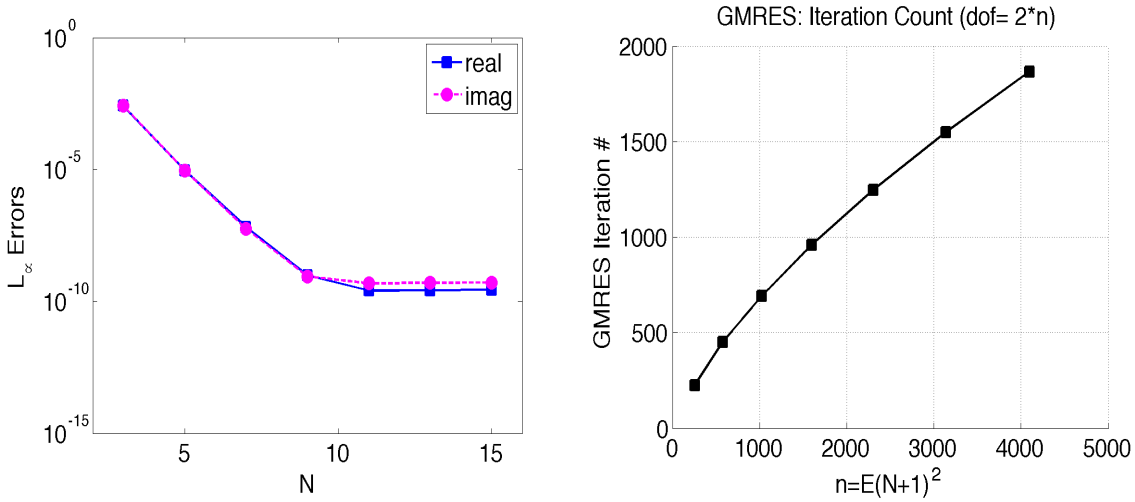


(c) Imaginary Part: $\alpha = 0.1$ (left) and $\alpha = 1.0$ (right)

Figure 3.7. Impedance on curve grating: $k = 1.5$ (yellow); $k = 2.5$ (blue); DtN (top/bottom).



(a) Curve Single Layer



(b) Curve Double Layer

Figure 3.8. Convergence, GMRES iteration counts, and mesh; E=4×4 and $N = 3, 5, 7, 9, 11, 13, 15$.

3.4.3 Nonsmooth Interfaces for Double Layers

For nonsmooth interface structures in double-layer media, we consider rectangular-shaped, isosceles triangle-shaped, and sawtooth-shaped gratings, provided with the convergence using the energy defect [7–9, 46].

Energy Defect: To diagnose the convergence of our algorithm we appeal to the well–established energy conservation measure. We point out that *outside* the grooves, i.e. in the domain

$$\Omega_0 := \{y > |g|_{L^\infty}\} \bigcup \{y < -|g|_{L^\infty}\},$$

the solutions $u^\pm$ can be expressed via the Rayleigh expansions

$$u^+(x,y) = \sum_{p=-\infty}^{\infty} B_p^+ e^{i\alpha_p x + i\beta_p^+ y}, \quad u^-(x,y) = \sum_{p=-\infty}^{\infty} B_p^- e^{i\alpha_p x - i\beta_p^- y}. \tag{3.4.11}$$

In the case of real wavenumbers $k^\pm$ there is a principle of conservation of energy [46] for the TE mode which can be expressed as

$$\sum_{p\in U^+} \beta_p^+ \left|B_p^+\right|^2 + \sum_{p\in U^-} \beta_p^- \left|B_p^-\right|^2 = \beta_0^+. \tag{3.4.12}$$

Defining the energy as

$$E^\pm(l) := \mathrm{Im}\left\{\frac{1}{L}\int_0^L \overline{u^\pm}(x,l)(\partial_y u^\pm(x,l))\, dx\right\}, \tag{3.4.13}$$

and substituting (3.4.11) into (3.4.13), we have

$$E^+(l_1) - E^-(l_2) = \sum_{p\in U^+} \beta_p^+ \left|B_p^+\right|^2 + \sum_{p\in U^-} \beta_p^- \left|B_p^-\right|^2 = \beta_0^+, \tag{3.4.14}$$

where $l_1 > |g|_{L^\infty}$ and $l_2 < -|g|_{L^\infty}$. Now we define the "energy defect" as the following to measure the errors of our numerical solutions.

$$\varepsilon := \left|1 - \frac{E^+(l_1) - E^-(l_2)}{\beta_0^+}\right|. \tag{3.4.15}$$

Double Layer: We consider 2π-periodicity in x for $(x,y) \in \Omega = [0,2\pi] \times [-1,1]$ including two different medium: Region 1= $[0,2\pi] \times [g(x),1]$ and Region 2=$[0,2\pi] \times$

$[-1, g(x)]$ where $g(x)$ is defined for the rectangular-shaped, isosceles triangle-shaped, and sawtooth-shaped interfaces as shown in Figures 3.4.3–3.4.3. We consider DtN boundaries at $y = -1$ and $y = 1$ and solve the numerical solution $\underline{\mathbf{u}}^N$ for the total field. We apply the incident field in the region $[0, 2\pi] \times [g(x), 1]$ defined by

$$u_{\mathrm{r}}^{\mathrm{inc}}(x, y) = \cos(\alpha x - \beta y) \tag{3.4.16}$$

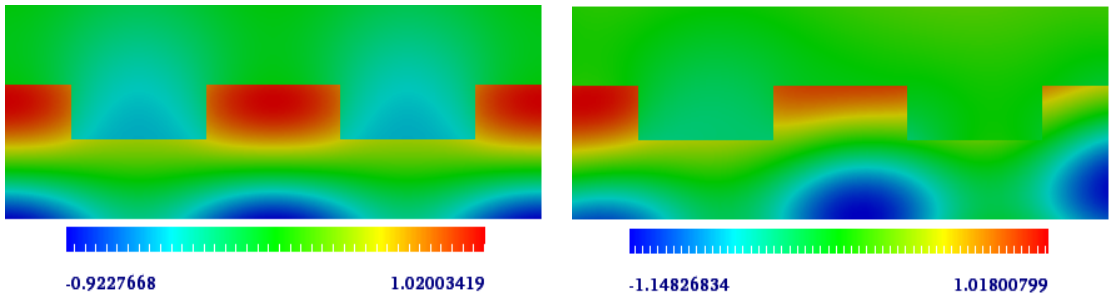
$$u_{\mathrm{i}}^{\mathrm{inc}}(x, y) = \sin(\alpha x - \beta y), \tag{3.4.17}$$

where $\beta = \sqrt{k^2 - \alpha^2}$. We examine the case of $k = k_1 = 1.5$ on $[0, 2\pi] \times [g(x), 1]$ and with $k = k_2 = 2.5$ on $[0, 2\pi] \times [-1, g(x)]$ for varying incident angle with $\alpha = 0.1$ and $\alpha = 1.0$.

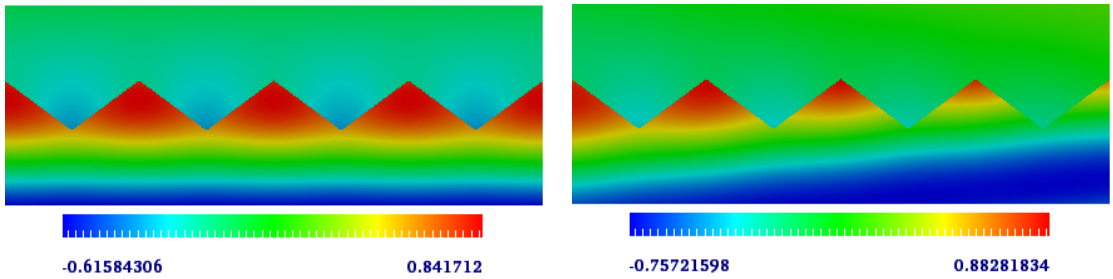
Figures 3.4.3–3.4.3 show our numerical solution for the scattered field. In Table 3.2, we demonstrate the convergence of our numerical solutions using the energy defect. They show spectral convergence as N increases with the GMRES interation counts increasing up to $700 \sim 1400$ for $N = 9$.

3.5 Conclusion

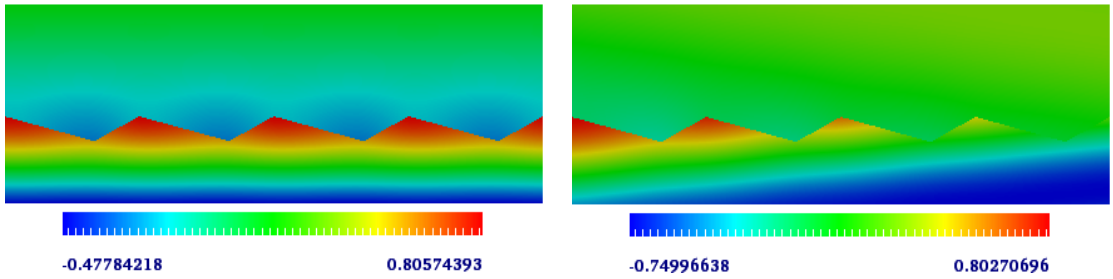
We have developed spectral element method in combination with the DtN boundary treatment for the Helmholtz operator describing acoustic scattering waves in single- and double-layer media. We consider example problems for the geometries with smooth flat structures, smooth curved structures, and nonsmooth interface structures. We solve the discretized spectral element scheme using GMRES interation technique. We validate our computational results provided with convergence studies in comparison with exact solutions, FTE solutions, and the energy defect measure, demonstrating exponential convergence. To be more efficient with reduced iterations, we consider developing preconditining technique based on a fast diagonalization method as the future work.



(a) Square: $\alpha = 0.0$ (left) and $\alpha = 0.2$ (right); $(E, N) = (64, 7)$

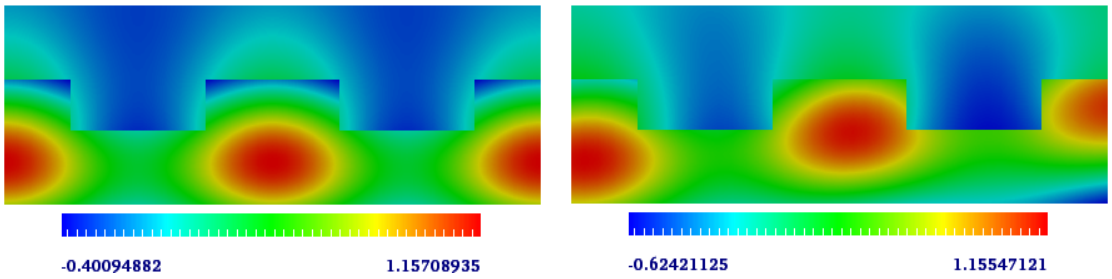


(b) Triangle: $\alpha = 0.0$ (left) and $\alpha = 0.2$ (right); (E,N)=(48,7)

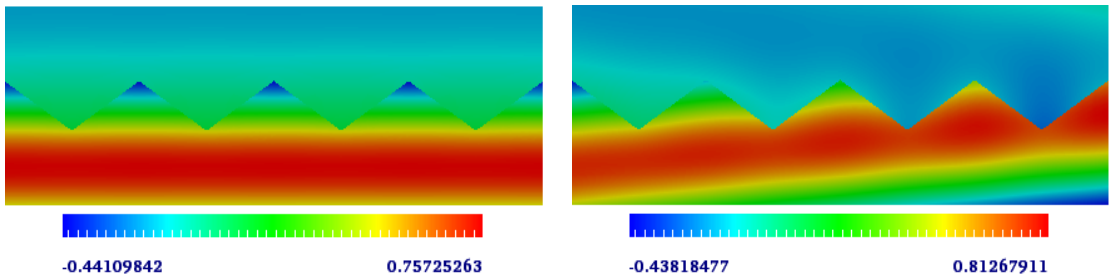


(c) Sawtooth 2: $\alpha = 0.0$ (left) and $\alpha = 0.2$ (right); $(E, N) = (48, 7)$

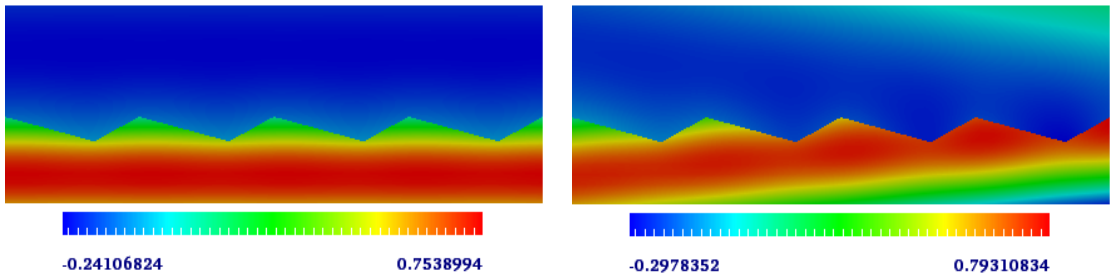
Figure 3.9. Scattered fields (Real part).



(a) Square: $\alpha = 0.0$ (left) and $\alpha = 0.2$ (right); $(E, N) = (64, 7)$



(b) Triangle: $\alpha = 0.0$ (left) and $\alpha = 0.2$ (right); $(E, N) = (48, 7)$



(c) Sawtooth 2: $\alpha = 0.0$ (left) and $\alpha = 0.2$ (right); $(E, N) = (48, 7)$

Figure 3.10. Scattered fields (Imaginary part).

Table 3.2
Convergence of the energy defect error ε.

Model A: Square							
Normal Incident				Oblique Incident			
E	N	ε	iter #	E	N	ε	iter #
64	3	0.435242E-03	226	64	3	0.429733E-03	309
	5	0.713931E-06	447		5	0.701378E-06	638
	7	0.496637E-09	704		7	0.488022E-09	1001
	9	0.793883E-12	998		9	0.139609E-11	1412
Model B: Triangle							
Normal Incident				Oblique Incident			
E	N	ε	iter #	E	N	ε	iter #
48	3	0.696253E-02	160	48	3	0.679565E-02	177
	5	0.481138E-04	321		5	0.470743E-04	349
	7	0.135466E-06	515		7	0.131819E-06	556
	9	0.208358E-09	728		9	0.187316E-09	782
Model C: Sawtooth							
Normal Incident				Oblique Incident			
E	N	ε	iter #	E	N	ε	iter #
48	3	0.140770E-01	182	48	3	0.136634E-01	186
	5	0.474567E-04	359		5	0.464225E-04	368
	7	0.133787E-06	563		7	0.130213E-06	574
	9	0.194617E-09	803		9	0.182961E-09	813

CHAPTER 4. A NEW SPECTRAL METHOD FOR NUMERICAL SOLUTION OF THE UNBOUNDED ROUGH SURFACE SCATTERING PROBLEM

In this chapter, a new spectral method is presented for solving the unbounded rough surface scattering problem, which is referred to as a non-local perturbation of an infinite plane surface such that the whole surface lies within a finite distance of the original plane. The method uses a transformed field expansion to reduce the boundary value problem with a complex scattering surface into a successive sequence of transmission problems of the Helmholtz equation with a plane surface. Hermite orthonormal basis functions are used to further simply the transmission problems to fully decoupled one-dimensional two-point boundary value problems with piecewise constant wavenumbers, which can be solved efficiently by a Legendre-Galerkin method. Numerical examples are presented for both the rough surface scattering and the plane surface scattering, where the analytic solution is available. Ample numerical results presented in the dissertation indicate that the new spectral method is efficient, accurate, and well suited to solve the scattering problem by unbounded rough surfaces.

4.1 Introduction

The phenomenon of acoustic and electromagnetic scattering by unbounded rough surfaces has received much attention from both the engineering and mathematical communities for its important applications in a wide range of scientific areas, such as modeling acoustic and electromagnetic wave propagation over outdoor ground and sea surfaces, optical scattering from the surface of materials in near-field optics or nano-optics, detection of underwater mines, especially those buried in soft sediments.

An unbounded rough surface is referred to as a non-local perturbation of an infinite plane surface such that the whole surface lies within a finite distance of the original plane. Due to the non-locally perturbed scattering surfaces, precise modeling and accurate computing present challenging mathematical and computational questions.

The uniqueness for the scattering by infinite rough surfaces was studied by Chandler-Wilde and Zhang [48]. The well-posedness of the rough surface scattering problems for the Helmholtz equation was investigated by Chandler-Wilde and Monk [49], Chanlder-Wilde et al. [50], and Lechleiter and Ritterbusch [51], Li and Shen [52], who considered variational approaches to solve a two- or three-dimensional rough surface scattering problem which models the time-harmonic acoustic wave scattering by a layer of homogeneous or inhomogeneous medium above a sound soft rough surface. In the work by Li et al. [53], the scattering problem was considered for the vector form of Maxwell's equations with dielectric surfaces, which models the time-harmonic electromagnetic wave by three layers of inhomogeneous medium with two infinite rough surfaces. In addition, the two-dimensional scalar model problem was considered by integral equation methods, where it assumes that the medium is homogeneous and the surface is the graph of a sufficiently smooth bounded function, when the boundary integral equation methods are applicable, e.g., Chandler-Wilde et al. [54, 55], Zhang and Chandler-Wilde [56, 57], and DeSanto and Martin [58–60]. We refer to Ritterbusch [61], Chandler-Wilde and Elschner [62] for related scattering problems where Sobolev spaces were studied for unbounded domains. Besides, a considerable amount of information is available for the solutions of the rough surface scattering problems by using approximate, asymptotic, or statistical methods, see, e.g., the reviews and monographs by Ogilvy [20], Voronovich [21], Saillard and Sentenac [22], Warnick and Chew [23], DeSanto [24], Elfouhaily and Guerin [25], and references cited therein. Despite the large amount of work done so far, we are not aware of any efficient and accurate numerical method for solving the scattering problem by unbounded rough surfaces.

The present work is concerned with the numerical solution for the unbounded rough surface scattering problem. Specifically, we study the acoustic wave propagation problem of the two-dimensional Helmholtz equation with an unbounded penetrable scattering surface. We consider the scattering of a time-harmonic wave field, generated from a point source, incident on an infinite rough surface from the top, where the spaces above and below the scattering surface are filled with some homogeneous absorbing materials, which accounts for the dielectric permittivity with positive imaginary part. The scattering phenomenon is modeled as a boundary value problem for acoustic wave propagation governed by the two-dimensional Helmholtz equation with transparent boundary conditions proposed on plane surfaces confining the scattering surface. Under the assumption that scattering rough surface is a sufficiently small and smooth deformation of a plane surface, we use the transformed field expansion to reduce the two-dimensional Helmholtz equation with complex scattering surface into a successive sequence of the transmission problems with a plane interface and piecewise constant wavenumbers. We use Hermite orthonormal basis functions, which plays the role of Fourier transform mathematically, to handle the difficulty from the infinite domain in horizontal direction, and further reduce the two-dimensional transmission problems into fully decoupled one-dimensional two-point boundary value problems, which are solved by a Legendre-Galerkin method. Numerical examples are considered for both the rough surface scattering and the plane surface scattering, where the analytic solution is available. Numerical errors are reported for the perturbation parameter and the wavenumbers, and for the truncation in the horizontal direction of the Hermite expansion, in the vertical direction of the Legendre expansion, and in transformed field power series expansion.

For boundary perturbation methods, we refer to a series of papers by Bruno and Reitich [63–65], Nicholls and Reitich [66], and references cited therein, for the rigorous mathematical and numerical analysis for solving some diffraction grating and obstacle scattering problems. An improved boundary perturbation algorithm, termed as transformed field expansion, was proposed by Nicholls and Reitich [67],

where a change of variables was done first to flatten the shape of the scattering surface and then followed by the boundary perturbation technique. The transformed field expansion method was shown to be accurate, stable, and robust even at high order, see, e.g., Nicholls and Shen [68] and Fang et al. [69] for solving the two- and three-dimensional bounded obstacle scattering problems. Recently, He et al. [70] developed an efficient and stable spectral method for the two-dimensional Helmholtz equation in a two-layered periodic structure, where a Legendre-Galerkin approximation was used to solve the reduced one-dimensional problems.

The outline of this chapter is as follows. In Section § 4.1, a mathematical model is introduced for the two-dimensional unbounded rough surface scattering problem which is formulated into a boundary value problem by using a transparent boundary condition. Section § 4.2 is devoted to the transformed field expansion which reduces the scattering problem of the two-dimensional Helmholtz equation with a complex scattering surface into a sequence of the transmission problems with a plane interface. In Section § 4.3, numerical approximations are considered for the transmission problems, where the Hermite orthonormal basis functions are used to decouple the two-dimensional problems into a sequence of one-dimensional problems which are then solved by a Legendre-Galerkin method. Numerical examples are presented to demonstrate the efficiency and accuracy of the proposed method in Section § 4.4. The paper is concluded with some general remarks and directions for future research in Section § 4.5.

4.2 Mathematical Model for Rough Surface Scattering

In this section, we shall introduce a mathematical model and define some notation for the scattering problem by an unbounded rough surface. Let the scattering surface be described by the curve $S = \{(x, y) : y = f(x), x \in \mathbb{R}\}$ with a bounded and Lipschitz continuous function f, as seen in Figure 6.1. The scattering surface S is

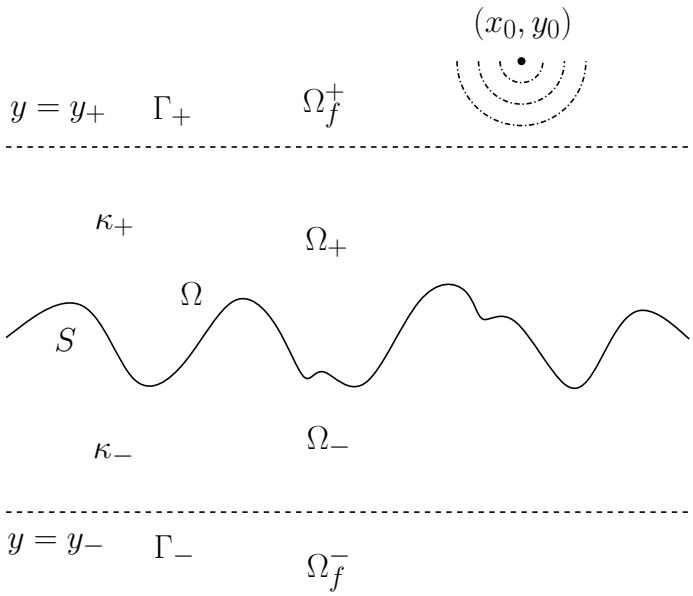


Figure 4.1. Problem geometry. A wave from the point source at (x_0, y_0) is incident on the scattering surface S from the top. The spaces Ω_f^+ (above S) and Ω_f^- (below S) are filled with materials whose wavenumbers are constants κ_+ and κ_-, respectively.

embedded in the strip

$$\Omega = \{(x, y) \in \mathbb{R}^2 : y_- < y < y_+\} = \mathbb{R} \times (y_-, y_+),$$

where y_- is a negative constant and y_+ is a positive constant. Let $\Omega_f^+ = \{(x, y) : y > f(x)\}$ and $\Omega_f^- = \{(x, y) : y < f(x)\}$ be filled with materials whose wavenumbers are constants κ_+ and κ_-, respectively. In fact, the wavenumber satisfies $\kappa_\pm^2 = \omega^2 \mu \varepsilon_\pm$, where ω is the angular frequency, μ is the magnetic permeability which is assumed to be a constant everywhere, and $\varepsilon_\pm$ is the electric permittivity in $\Omega_f^\pm$. In this work, the electric permittivity $\varepsilon_\pm$ are assumed to be two complex numbers with positive imaginary parts. The condition $\mathrm{Im}\kappa_\pm^2 > 0$ physically accounts for energy absorption and mathematically ensures the existence and uniqueness of the solution. We also denote by $\Gamma_+ = \{y = y_+\}$ and $\Gamma_- = \{y = y_-\}$ the top and bottom boundaries of the domain Ω.

Suppose that a wave generated from a point source is incident on S from the top. Explicitly, the point incident field is taken as the fundamental solution of the Helmholtz equation in Ω_+, i.e.,

$$u^{\mathrm{inc}}(x, y; x_0, y_0) = \frac{\mathrm{i}}{4} H_0^{(1)} \left(\kappa_+ |(x, y) - (x_0, y_0)|\right), \tag{4.2.1}$$

where $H_0^{(1)}$ is the Hankel function of first kind with order zero, (x, y) is the observation point, and (x_0, y_0) is a given source point in Ω_+. Clearly the incident field satisfies the two dimensional Helmholtz equation:

$$\Delta u^{\mathrm{inc}}(x, y) + \kappa_+^2 u^{\mathrm{inc}}(x, y) = -\delta(x - x_0)\delta(y - y_0) \quad \text{in } \mathbb{R}^2,$$

where δ is the Dirac delta function.

The scattering of time harmonic electromagnetic waves in the transverse electric case can also be modeled by the two dimensional Helmholtz equation:

$$\Delta u(x, y) + \kappa^2 u(x, y) = -\delta(x - x_0)\delta(y - y_0) \quad \text{in } \mathbb{R}^2, \tag{4.2.2}$$

where the wavenumber

$$\kappa = \begin{cases} \kappa_+ & \text{in } \Omega_f^+, \\ \kappa_- & \text{in } \Omega_f^-. \end{cases}$$

Due to the unbounded scattering surface, the usual Sommerfeld radiation condition is no longer valid. The radiation condition that we impose is the boundedness of u as y tends to infinity. More precisely, we insist that u is composed of bounded outgoing waves in Ω_+ and Ω_- plus the incident wave u^{inc} in Ω_+.

For any given u on $\Gamma_\pm$, define the boundary operators $T_\pm$:

$$T_\pm u = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \pm \mathrm{i}\beta_\pm \hat{u}(\xi, y_\pm) e^{\mathrm{i}\xi x} d\xi,$$

where

$$\beta_\pm^2(\xi) = \kappa_\pm^2 - |\xi|^2 \quad \text{with } \mathrm{Im}\beta_\pm(\xi) > 0.$$

Following [52], we can deduce a transparent boundary condition on $\Gamma_\pm$:

$$\partial_y u = T_\pm u + \rho^\pm \quad \text{on } \Gamma_\pm, \tag{4.2.3}$$

where

$$\rho^+ = \partial_y u^{\text{inc}} - T_+ u^{\text{inc}} \quad \text{and} \quad \rho^- = 0. \tag{4.2.4}$$

Next we reformulate the scattering problem (4.2.2) and (4.2.3) into a transmission problem, to which we will apply the transformed field expansion [67]. Denote $\Omega_\pm = \Omega_f^\pm \cap \Omega$, as seen in Figure 6.1. Consider the Helmholtz equation (4.2.2) in $\Omega_\pm$:

$$\Delta u^\pm + \kappa_\pm^2 u^\pm = 0 \quad \text{in } \Omega_\pm. \tag{4.2.5}$$

Recall the non-local transparent boundary conditions (4.2.3)

$$\partial_y u^\pm = T_\pm u^\pm + \rho^\pm \quad \text{on } \Gamma_\pm. \tag{4.2.6}$$

Following from the jump conditions, we obtain that the field and its normal derivative are continuous across the scattering surface S, i.e.,

$$u^+(x, f(x)) = u^-(x, f(x)), \tag{4.2.7}$$

$$\partial_{\mathbf{n}} u^+(x, f(x)) = \partial_{\mathbf{n}} u^-(x, f(x)), \tag{4.2.8}$$

where $\mathbf{n} = (n_1, n_2)^\top$ is the unit normal vector pointing from Ω_+ to Ω_-. Explicitly, we have

$$n_1 = \frac{f'(x)}{\sqrt{1 + [f'(x)]^2}} \quad \text{and} \quad n_2 = -\frac{1}{\sqrt{1 + [f'(x)]^2}}.$$

Hence, the transmission problem is to find the fields u^+ and u^-, which satisfy the Helmholtz equation (4.2.5), the boundary condition (4.2.6), and the continuity conditions (4.2.7) and (4.2.8). It was shown in Li and Shen [52] that the transmission problem has a unique weak solution; furthermore, an analytic solution as an infinite series is deduced under the assumption that the scattering surface S is a sufficiently small and smooth deformation of a plane surface.

4.3 Transformed Field Expansion

The transformed field expansion method, as applied to the unbounded rough surface scattering, begins with the change of variables:

$$x_1 = x, \quad y_1 = y_+ \left(\frac{y - f}{y_+ - f} \right), \quad f < y < y_+,$$

and

$$x_2 = x, \quad y_2 = y_- \left(\frac{y - f}{y_- - f} \right), \quad y_- < y < f,$$

which maps the perturbed domains Ω_+ and Ω_- to unperturbed strip domains $D_+ = \{(x, y) \in \mathbb{R}^2 : 0 < y < y_+\}$ and $D_- = \{(x, y) \in \mathbb{R}^2 : y_- < y < 0\}$, respectively. We now seek to restate the transmission problem (4.2.5)–(4.2.8) in these transformed coordinates. It is easy to verify the differentiation rules

$$\partial_x = \partial_{x_1} - f' \left(\frac{y_+ - y_1}{y_+ - f} \right) \partial_{y_1},$$
$$\partial_y = \left(\frac{y_+}{y_+ - f} \right) \partial_{y_1},$$

for $f < y < y_+$, and

$$\partial_x = \partial_{x_2} - f' \left(\frac{y_- - y_2}{y_- - f} \right) \partial_{y_2},$$
$$\partial_y = \left(\frac{y_-}{y_- - f} \right) \partial_{y_2},$$

for $y_- < y < f$.

Introduce new functions $w^+(x_1, y_1) = u^+(x, y)$ and $w^-(x_2, y_2) = u^-(x, y)$ under the transformation. It can be verified after tedious but straightforward calculations that $w^\pm$, upon dropping the subscript, satisfy the equation

$$c_1^\pm \frac{\partial^2 w^\pm}{\partial x^2} + c_2^\pm \frac{\partial^2 w^\pm}{\partial y^2} + c_3^\pm \frac{\partial^2 w^\pm}{\partial x \partial y} + c_4^\pm \frac{\partial w^\pm}{\partial y} + c_1^\pm \kappa_\pm^2 w^\pm = 0 \quad \text{in } D_\pm, \tag{4.3.1}$$

where

$$\begin{aligned} c_1^\pm &= (y_\pm - f)^2, \\ c_2^\pm &= [f'(y_\pm - y)]^2 + y_+^2, \\ c_3^\pm &= -2f'(y_\pm - y)(y_\pm - f), \\ c_4^\pm &= -(y_\pm - y)[f''(y_\pm - f) + 2(f')^2]. \end{aligned}$$

The non-local transparent boundary conditions (4.2.6) are

$$\partial_y w^\pm = \left(1 - \frac{f}{y_\pm}\right)\left(T_\pm w^\pm + \rho^\pm\right) \quad \text{on } \Gamma_\pm. \tag{4.3.2}$$

The continuity conditions (4.2.7) and (4.2.8) reduce to

$$w^+(x, 0) = w^-(x, 0), \tag{4.3.3}$$

$$\left(\frac{y_+}{y_+ - f}\right) \partial_y w^+(x, 0) = \left(\frac{y_-}{y_- - f}\right) \partial_y w^-(x, 0). \tag{4.3.4}$$

Now, we use a classical boundary perturbation argument. We assume that the scattering surface S is a sufficiently small perturbation of the flat plane, i.e., $f = \varepsilon g$ with ε sufficiently small. We consider the formal expansions of $w^\pm$ in a power series of ε:

$$w^\pm(x, y; \varepsilon) = \sum_{k=0}^{\infty} w_k^\pm(x, y)\varepsilon^k. \tag{4.3.5}$$

Substituting $f = \varepsilon g$ into $c_j^\pm$ and inserting the above expansions into (4.3.1), we may derive the recursions for $w_k^\pm$:

$$\frac{\partial^2 w_k^\pm}{\partial x^2} + \frac{\partial^2 w_k^\pm}{\partial y^2} + \kappa_\pm^2 w_k^\pm = v_k^\pm \quad \text{in } D_\pm, \tag{4.3.6}$$

where

$$\begin{aligned} v_k^{\pm} =& \frac{2g}{y_\pm}\frac{\partial^2 w_{k-1}^{\pm}}{\partial x^2} + \frac{2g'(y_\pm - y)}{y_\pm}\frac{\partial^2 w_{k-1}^{\pm}}{\partial x \partial y} + \frac{g''(y_\pm - y)}{y_\pm}\frac{\partial w_{k-1}^{\pm}}{\partial y} + \frac{2\kappa_\pm^2 g}{y_\pm} w_{k-1}^{\pm} \\ & - \frac{g^2}{y_\pm^2}\frac{\partial^2 w_{k-2}^{\pm}}{\partial x^2} - \frac{(g')^2(y_\pm - y)^2}{y_\pm^2}\frac{\partial^2 w_{k-2}^{\pm}}{\partial y^2} - \frac{2gg'(y_\pm - y)}{y_\pm^2}\frac{\partial^2 w_{k-2}^{\pm}}{\partial x \partial y} \\ & + \frac{[2(g')^2 - gg''](y_\pm - y)}{y_\pm^2}\frac{\partial w_{k-2}^{\pm}}{\partial y} - \frac{\kappa_\pm^2 g^2}{y_\pm^2} w_{k-2}^{\pm}. \end{aligned}$$

The non-local boundary conditions (4.3.2) become

$$\partial_y w_k^{\pm} - T_\pm w_k^{\pm} = \rho_k^{\pm}, \quad y = y_\pm, \tag{4.3.7}$$

where

$$\rho_0^+ = \rho, \quad \rho_1^+ = -\left(\frac{g}{y_+}\right) T_+ w_0^+ - \left(\frac{g}{y_+}\right)\rho, \quad \rho_k^+ = -\left(\frac{g}{y_+}\right) T_+ w_{k-1}^+, \quad k = 2, 3, \ldots,$$
$$\rho_0^- = 0, \quad \rho_k^- = \left(\frac{g}{y_-}\right) T_- w_{k-1}^-, \quad k = 1, 2, \ldots.$$

The continuity conditions (4.3.3) and (4.3.4) at the interface $y = 0$ reduce to

$$w_k^+(x, 0) - w_k^-(x, 0) = 0, \tag{4.3.8}$$

$$\partial_y w_k^+(x, 0) - \partial_y w_k^-(x, 0) = h_k, \tag{4.3.9}$$

where

$$h_0 = 0, \quad h_k = \left(\frac{g}{y_-}\right)\partial_y w_{k-1}^+ - \left(\frac{g}{y_+}\right)\partial_y w_{k-1}^-, \quad k = 1, 2, \ldots.$$

Note that the Helmholtz problem (4.3.6) for the current terms $w_k^\pm$ involve some non-homogeneous terms $v_k^\pm$, $\rho_k^\pm$, and h_k, which only depend on previous two terms $w_{k-1}^\pm$ and $w_{k-2}^\pm$. Thus, the transmission problem (4.3.6)–(4.3.9) in rectangular domains $D_\pm$ indeed can be solved efficiently in a recursive manner starting from $k = 0$.

A main difficulty in numerically solving the transmission problem (4.3.6)–(4.3.9) is how to treat the non-local boundary conditions in (4.3.7). It is shown in [52] that by taking the Fourier transform in x, the boundary conditions (4.3.7) become *local* in the Fourier frequency space. Indeed, dropping the subscript k for simplicity of

notation, and taking the Fourier transform of (4.3.6) with respect to the variable x, we obtain

$$\frac{\partial^2 \hat{w}^{\pm}}{\partial y^2} + (\kappa_{\pm}^2 - \xi^2)\hat{w}^{\pm} = \hat{v}^{\pm}. \tag{4.3.10}$$

The non-local boundary conditions (4.3.7) become:

$$\partial_y \hat{w}^{\pm} \mp \mathrm{i}\beta_{\pm}\hat{w}^{\pm} = \hat{\rho}^{\pm}, \tag{4.3.11}$$

which is local in the Fourier variable ξ. The continuity conditions reduce to

$$\hat{w}^{+}(\xi, 0) - \hat{w}^{-}(\xi, 0) = 0, \tag{4.3.12}$$

$$\partial_y \hat{w}^{+}(\xi, 0) - \partial_y \hat{w}^{-}(\xi, 0) = \hat{h}. \tag{4.3.13}$$

We observe that for each $\xi \in \mathbb{R}$, the problem (4.3.10)–(4.3.13) is an one-dimensional two-point boundary value problem whose solution can expressed analytically [52]. However, these analytic expressions are of limited use in practice, since the solution is expressed in the Fourier variable ξ which can not readily converted to the physical variable x due to the lack of discrete Fourier transform in $\mathbb{R}$. In the next section, we use orthonormal Hermit basis functions which play the role of the Fourier transform numerically and allow us to reduce the two-dimensional problem (4.3.6)–(4.3.9) into a sequence of one-dimensional problems that can be solved efficiently and accurately by a Legendre-Galerkin method.

4.4 Approximation by Hermit Functions

In this section, we consider the approximation to the transmission problem (4.3.10) – (4.3.13). We use the Hermite orthonormal functions for the horizontal x-direction, and the Legendre-Galerkin method for the reduced one-dimensional problem in the vertical y-direction.

4.4.1 Hermite Orthonormal Basis

Denote by $H_m(x)$ the Hermite polynomial of degree m on $\mathbb{R}$ for $m = 0, 1, 2, \cdots$. These polynomials are orthogonal with respect to the weigh function e^{-x^2}, i.e.,

$$\int_{\mathbb{R}} H_m(x) H_k(x) e^{-x^2} dx = 2^m m! \sqrt{\pi}\, \delta_{mk}, \tag{4.4.1}$$

where δ_{mk} is the Dirac delta function. The sequence of Hermite polynomials satisfies the recursion

$$H_{m+1}(x) = 2xH_m(x) - 2mH_{m-1}(x) \tag{4.4.2}$$

and the identity

$$H'_m(x) = 2mH_{m-1}(x). \tag{4.4.3}$$

Define a sequence of Hermite functions

$$\psi_m(x) = \left(2^m m! \sqrt{\pi}\right)^{-1/2} H_m(x) e^{-x^2/2}.$$

It follows from (4.4.1) that the Hermite functions form an orthonormal basis of the Hilbert space $L^2(\mathbb{R})$, i.e.,

$$\int_{\mathbb{R}} \psi_m(x) \psi_k(x) dx = \delta_{mk}. \tag{4.4.4}$$

It follows from (4.4.2) and (4.4.3) that the Hermite functions satisfy the recursion

$$\psi_{m+1}(x) = \sqrt{\frac{2}{m+1}}\, x\psi_m(x) - \sqrt{\frac{m}{m+1}}\, \psi_{m-1}(x) \tag{4.4.5}$$

and the identity

$$\psi'_m(x) = -x\psi_m(x) + \sqrt{2m}\, \psi_{m-1}(x). \tag{4.4.6}$$

The following result plays a key role in our algorithm presented below. We refer to Duoandikoetxea [71] (cf. page 22) for the proof.

Lemma 4.4.1 The Hermite function ψ_m is the eigenfunction of the Fourier transform operator with eigenvalue $(-\mathrm{i})^m, m = 0, 1, \ldots$, which means

$$\hat{\psi}_m(\xi) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \psi_m(x) e^{-\mathrm{i}\xi x} dx = (-\mathrm{i})^m \psi_m(\xi).$$

Using the Hermite basis functions, we consider the following expansions

$$w^{\pm}(x,y) = \sum_{m=0}^{\infty} w_m^{\pm}(y)\psi_m(x),$$
$$v^{\pm}(x,y) = \sum_{m=0}^{\infty} v_m^{\pm}(y)\psi_m(x),$$
$$\rho^{\pm}(x) = \sum_{m=0}^{\infty} \rho_m^{\pm}\psi_m(x),$$
$$h(x) = \sum_{m=0}^{\infty} h_m\psi_m(x).$$

Taking the Fourier transform with respect to x of above expansions and using Lemma 4.4.1, we have

$$\hat{w}^{\pm}(\xi,y) = \sum_{m=0}^{\infty} (-\mathrm{i})^m w_m^{\pm}(y)\psi_m(\xi),$$
$$\hat{v}^{\pm}(\xi,y) = \sum_{m=0}^{\infty} (-\mathrm{i})^m v_m^{\pm}(y)\psi_m(\xi),$$
$$\hat{\rho}^{\pm}(\xi) = \sum_{m=0}^{\infty} (-\mathrm{i})^m \rho_m^{\pm}\psi_m(\xi),$$
$$\hat{h}(\xi) = \sum_{m=0}^{\infty} (-\mathrm{i})^m h_m\psi_m(\xi).$$

Plugging the above expansions into (4.3.10) yields

$$\sum_{m=0}^{\infty}\left[(-\mathrm{i})^m \frac{\mathrm{d}^2 w_m^{\pm}(y)}{\mathrm{d}y^2}\psi_m(\xi) + (-\mathrm{i})^m(\kappa_{\pm}^2 - \xi^2)w_m^{\pm}(y)\psi_m(\xi)\right] = \sum_{m=0}^{\infty}(-\mathrm{i})^m v_m^{\pm}(y)\psi_m(\xi). \tag{4.4.7}$$

The boundary condition (4.3.11) reduce to

$$\sum_{m=0}^{\infty}\left[(-\mathrm{i})^m \frac{\mathrm{d}w_m^{\pm}(y_{\pm})}{\mathrm{d}y}\psi_m(\xi) \mp (-\mathrm{i})^m \mathrm{i}\beta_{\pm} w_m^{\pm}(y_{\pm})\psi_m(\xi)\right] = \sum_{m=0}^{\infty}(-\mathrm{i})^m \rho_m^{\pm}\psi_m(\xi). \tag{4.4.8}$$

The continuity conditions (4.3.12) and (4.3.13) reduce to

$$\sum_{m=0}^{\infty}(-\mathrm{i})^m \left[w_m^{+}(0) - w_m^{-}(0)\right]\psi_m(\xi) = 0, \tag{4.4.9}$$

and

$$\sum_{m=0}^{\infty}(-\mathrm{i})^m \left[\frac{\mathrm{d}w_m^{+}(0)}{\mathrm{d}y} - \frac{\mathrm{d}w_m^{-}(0)}{\mathrm{d}y}\right]\psi_m(\xi) = \sum_{m=0}^{\infty}(-\mathrm{i})^m h_m\psi_m(\xi). \tag{4.4.10}$$

Define a diagonal matrix $D = \mathrm{diag}((-\mathrm{i})^0, (-\mathrm{i})^1, (-\mathrm{i})^2, \dots, (-\mathrm{i})^m, \dots)$ and vectors

$$\begin{aligned}
\mathbf{w}^{\pm}(y) &= D \cdot (w_0^{\pm}(y), w_1^{\pm}(y), \dots, w_m^{\pm}(y), \dots)^{\top}, \\
\mathbf{v}^{\pm}(y) &= D \cdot (v_0^{\pm}(y), v_1^{\pm}(y), \dots, v_m^{\pm}(y), \dots)^{\top}, \\
\boldsymbol{\rho}^{\pm} &= D \cdot (\rho_0^{\pm}, \rho_1^{\pm}, \dots, \rho_m^{\pm}, \dots)^{\top}, \\
\mathbf{h} &= D \cdot (h_1, h_2, \dots, h_m, \dots)^{\top}.
\end{aligned}$$

We also define matrices $A = (a_{ij})$ and $S^{\pm} = (s_{ij}^{\pm})$ with entries given by

$$a_{ij} = \int_{\mathbb{R}} \xi^2 \psi_j(\xi) \psi_i(\xi) d\xi.$$

and

$$s_{ij}^{\pm} = \int_{\mathbb{R}} \pm \mathrm{i}\beta_{\pm} \psi_j(\xi) \psi_i(\xi) d\xi.$$

In fact, it follows from (4.4.4) and (4.4.5) that the matrix A is symmetric and has only three nonzero diagonals with entries given by:

$$a_{ij} = \frac{\sqrt{(i+1)(j+1)} + \sqrt{ij}}{2} \delta_{i,j} + \frac{\sqrt{(i+1)j}}{2} \delta_{i+1,j-1} + \frac{\sqrt{i(j+1)}}{2} \delta_{i-1,j+1}.$$

Multiplying the Hermite basis function and integrating in $\mathbb{R}$, we may rewrite (4.4.7) into the matrix form:

$$\frac{\mathrm{d}^2 \mathbf{w}^{\pm}(y)}{\mathrm{d}y^2} + (\kappa_{\pm}^2 I - A)\mathbf{w}^{\pm}(y) = \mathbf{v}^{\pm}(y), \tag{4.4.11}$$

where I is the identity matrix. The boundary conditions (6.1.11) can be written as

$$\frac{\mathrm{d}\mathbf{w}^{\pm}(y_{\pm})}{\mathrm{d}y} - S^{\pm}\mathbf{w}^{\pm}(y_{\pm}) = \boldsymbol{\rho}^{\pm}. \tag{4.4.12}$$

The continuity conditions (4.4.9) and (4.4.10) can be written as

$$\mathbf{w}^{+}(0) - \mathbf{w}^{-}(0) = 0, \tag{4.4.13}$$

and

$$\frac{\mathrm{d}\mathbf{w}^{+}(0)}{\mathrm{d}y} - \frac{\mathrm{d}\mathbf{w}^{-}(0)}{\mathrm{d}y} = \mathbf{h}. \tag{4.4.14}$$

Therefore, we have reformulated (4.3.10)–(4.3.13) as a *coupled* infinite system (4.4.11)-(4.4.14). We shall prove below that A and $S^{\pm}$ commute, namely

$$AS^{\pm} = S^{\pm}A. \tag{4.4.15}$$

Therefore, A and $S^{\pm}$ can be simultaneously diagonalized, and consequently, the system (4.4.11)–(4.4.14) can be *decoupled*.

We now prove (4.4.15) through a sequence of lemmas below.

Lemma 4.4.2 For any real polynomials $g_1(\xi), g_2(\xi)$, and integers $i, j \geq 0$, it holds the identity

$$\sum_{k=0}^{\infty} \int_{\mathbb{R}} \int_{\mathbb{R}} g_1(\xi) g_2(\eta) \psi_i(\xi) \psi_k(\xi) \psi_k(\eta) \psi_j(\eta) d\xi d\eta = \int_{\mathbb{R}} g_1(\xi) g_2(\xi) \psi_i(\xi) \psi_j(\xi) d\xi. \tag{4.4.16}$$

Proof We prove this lemma by the method of induction. First, without loss of generality, we may assume g_1 is a constant function, e.g., $g_1 = 1$. By the orthogonality of the Hermite basis functions, we have

$$\begin{aligned}
&\sum_{k=0}^{\infty} \int_{\mathbb{R}} \int_{\mathbb{R}} g_1(\xi) g_2(\eta) \psi_i(\xi) \psi_k(\xi) \psi_k(\eta) \psi_j(\eta) d\xi d\eta \\
&\quad = \sum_{k=0}^{\infty} \int_{\mathbb{R}} g_2(\eta) \psi_k(\eta) \psi_j(\eta) d\eta \int_{\mathbb{R}} \psi_i(\xi) \psi_k(\xi) d\xi = \int_{\mathbb{R}} g_1(\eta) g_2(\eta) \psi_i(\eta) \psi_j(\eta) d\eta.
\end{aligned}$$

Next we show the same property holds for $g_1(\xi) = \xi$. Using the recursion (4.4.5) and the orthogonality of the Hermite basis functions yield

$$\begin{aligned}
&\sum_{k=0}^{\infty} \int_{\mathbb{R}} \int_{\mathbb{R}} g_1(\xi) g_2(\eta) \psi_i(\xi) \psi_k(\xi) \psi_k(\eta) \psi_j(\eta) d\xi d\eta \\
&= \sum_{k=0}^{\infty} \int_{\mathbb{R}} g_2(\eta) \psi_k(\eta) \psi_j(\eta) d\eta \int_{\mathbb{R}} \xi \psi_i(\xi) \psi_k(\xi) d\xi \\
&= \sum_{k=0}^{\infty} \int_{\mathbb{R}} g_2(\eta) \psi_k(\eta) \psi_j(\eta) d\eta \int_{\mathbb{R}} \left(\sqrt{\frac{i+1}{2}} \psi_{i+1}(\xi) + \sqrt{\frac{i}{2}} \psi_{i-1}(\xi) \right) \psi_k(\xi) d\xi \\
&= \int_{\mathbb{R}} g_2(\eta) \left(\sqrt{\frac{i+1}{2}} \psi_{i+1}(\eta) + \sqrt{\frac{i}{2}} \psi_{i-1}(\eta) \right) \psi_j(\eta) d\eta \\
&= \int_{\mathbb{R}} g_2(\eta) \eta \psi_i(\eta) \psi_j(\eta) d\eta = \int_{\mathbb{R}} g_1(\eta) g_2(\eta) \psi_i(\eta) \psi_j(\eta) d\eta.
\end{aligned}$$

Now we may assume that (4.4.16) holds for any polynomial $g_1(\xi)$ with degree less than or equal to n, we need to show that it also holds for any polynomial $g_1(\xi)$ with degree $n+1$. Here we can assume $g_1(\xi)$ has no constant term, as (4.4.16) is proved for any constant function of g_1. Then we can write $g_1(\xi) = \xi f(\xi)$, where $f(\xi)$ is a polynomial of degree n.

Using the recursion (4.4.5) and the orthogonality of the Hermite basis functions again, we have

$$\begin{aligned}
&\sum_{k=0}^{\infty} \int_{\mathbb{R}} \int_{\mathbb{R}} g_1(\xi) g_2(\eta) \psi_i(\xi) \psi_k(\xi) \psi_k(\eta) \psi_j(\eta) d\xi d\eta \\
&= \sum_{k=0}^{\infty} \int_{\mathbb{R}} g_2(\eta) \psi_k(\eta) \psi_j(\eta) d\eta \int_{\mathbb{R}} f(\xi) \xi \psi_i(\xi) \psi_k(\xi) d\xi \\
&= \sum_{k=0}^{\infty} \int_{\mathbb{R}} g_2(\eta) \psi_k(\eta) \psi_j(\eta) d\eta \int_{\mathbb{R}} f(\xi) \left(\sqrt{\frac{i+1}{2}} \psi_{i+1}(\xi) + \sqrt{\frac{i}{2}} \psi_{i-1}(\xi) \right) \psi_k(\xi) d\xi \\
&= \sum_{k=0}^{\infty} \sqrt{\frac{i+1}{2}} \int_{\mathbb{R}} \int_{\mathbb{R}} f(\eta) g_2(\xi) \psi_{i+1}(\xi) \psi_k(\xi) \psi_k(\eta) \psi_j(\eta) d\xi d\eta \\
&\qquad + \sum_{k=0}^{\infty} \sqrt{\frac{i}{2}} \int_{\mathbb{R}} \int_{\mathbb{R}} f(\eta) g_2(\xi) \psi_{i-1}(\xi) \psi_k(\xi) \psi_k(\eta) \psi_j(\eta) d\xi d\eta \\
&= \sqrt{\frac{i+1}{2}} \int_{\mathbb{R}} f(\eta) g_2(\eta) \psi_{i+1}(\eta) \psi_j(\eta) d\eta + \sqrt{\frac{i}{2}} \int_{\mathbb{R}} f(\eta) g_2(\eta) \psi_{i-1}(\eta) \psi_j(\eta) d\eta \\
&= \int_{\mathbb{R}} g_2(\eta) f(\eta) \left(\sqrt{\frac{i+1}{2}} \psi_{i+1}(\eta) + \sqrt{\frac{i}{2}} \psi_{i-1}(\eta) \right) \psi_j(\eta) d\eta \\
&= \int_{\mathbb{R}} g_2(\eta) f(\eta) \eta \psi_i(\eta) \psi_j(\eta) d\eta = \int_{\mathbb{R}} g_2(\eta) g_1(\eta) \psi_i(\eta) \psi_j(\eta) d\eta.
\end{aligned}$$

It follows from the method of induction that (4.4.16) holds for any polynomials.

Denote by $\langle \cdot, \cdot \rangle$ and $(\cdot, \cdot)$ the inner products in l^2 and L^2, respectively. Let $\mathbf{u} = [u_0, u_1, \dots]^\top, \mathbf{v} = [v_0, v_1, \dots]^\top \in l^2$, and

$$u(x) = \sum_{k=0}^{\infty} u_k \psi_k(x) \quad \text{and} \quad v(x) = \sum_{k=0}^{\infty} u_k \psi_k(x).$$

We have the following useful lemma.

Lemma 4.4.3 For any integer $k \geq 0$, it holds

$$\langle A^k \mathbf{u}, \mathbf{v} \rangle = (x^{2k} u(x), v(x)). \tag{4.4.17}$$

Proof We prove this lemma by the method of induction. Clearly, it follows from the definitions of the inner products in l^2 and L^2 that the identity (4.4.17) is satisfied, i.e.,

$$\langle \mathbf{u},\, \mathbf{v}\rangle = (u(x),\, v(x)).$$

So, we first prove the identity (4.4.17) is satisfied for $k = 1$. It follows from the definitions that we have

$$\begin{aligned}(x^2u(x), v(x)) &= \left(x^2 \sum_{i=0}^{\infty} u_i\psi_i(x),\, \sum_{j=0}^{\infty} v_j\psi_j(x)\right) = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} u_i\, v_j (x^2\psi_i(x),\, \psi_j(x)) \\ &= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} u_i\, v_j\, a_{ji} = \langle A\mathbf{u},\, \mathbf{v}\rangle.\end{aligned}$$

We assume that (4.4.17) is satisfied for some integer $k > 1$, i.e.,

$$\langle (A)^k\mathbf{u},\, \mathbf{v}\rangle = (x^{2k}u(x),\, v(x)).$$

Next we show that (4.4.17) is satisfied for the integer $k+1$. By definitions and Lemma 4.4.2 , we have

$$\begin{aligned}&(x^{2(k+1)}u(x),\, v(x)) = (x^2(x^{2k}u(x)),\, v(x)) = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} u_i\, v_j (x^2x^{2k}\psi_i(x),\, \psi_j(x)) \\ &= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} u_i\, v_j \sum_{m=0}^{\infty} (x^2\psi_i(x),\, \psi_m(x))(x^{2k}\psi_m(x),\, \psi_j(x)) \\ &= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} u_i\, v_j \sum_{m=0}^{\infty} a_{mi}\langle A^k\mathbf{e}_m,\, \mathbf{e}_j\rangle = \sum_{m=0}^{\infty}\left(\sum_{i=0}^{\infty} u_i\, a_{mi}\right)\left(\sum_{j=0}^{\infty} v_j\langle A^k\mathbf{e}_m,\, \mathbf{e}_j\rangle\right) \\ &= \sum_{m=0}^{\infty} (A\mathbf{u})_m\, \langle A^k\mathbf{e}_m,\, \mathbf{v}\rangle = \langle A^k \sum_{m=0}^{\infty} (A\mathbf{u})_m\mathbf{e}_m,\, \mathbf{v}\rangle = \langle A^k\, A\mathbf{u},\, \mathbf{v}\rangle = \langle A^{k+1}\,\mathbf{u},\, \mathbf{v}\rangle,\end{aligned}$$

where $\mathbf{e}_m$ is the unit vector, whose m^{th} component is one and others are zeros. Therefore, the identity (4.4.17) is satisfied for any integer $k \geq 0$.

Lemma 4.4.4 The matrices $S^{\pm}$ can be expressed as

$$S^{\pm} = \sum_{k=0}^{\infty} \frac{g_{\pm}^{(k)}(0)}{k!}\, A^k, \tag{4.4.18}$$

where $g_{\pm}(\xi) = \mathrm{i}\sqrt{{\kappa_{\pm}}^2 - \xi}$.

Proof Since $\kappa_\pm$ are given complex numbers and $g_\pm(\xi)$ are analytic on the whole real axis, we have the Maclaurin series expansions for $g_\pm(\xi)$:

$$g_\pm(\xi) = \sum_{k=0}^{\infty} \frac{g_\pm^{(k)}(0)}{n!}\xi^k \quad \text{for all real } \xi. \tag{4.4.19}$$

It follows from (4.4.17) that we have

$$\begin{aligned}
\langle S^\pm \mathbf{e}_j,\, \mathbf{e}_i\rangle &= (g_\pm(\xi^2)\psi_j(\xi),\, \psi_i(\xi)) = \left(\sum_{k=0}^{\infty} \frac{g_\pm^{(k)}(0)}{k!}(\xi^{2k})\psi_j(\xi),\, \psi_i(\xi)\right), \\
&= \sum_{k=0}^{\infty} \frac{g_\pm^{(k)}(0)}{k!}(\xi^{2k}\psi_j(\xi),\, \psi_i(\xi)) = \sum_{k=0}^{\infty} \frac{g_\pm^{(k)}(0)}{k!}\langle A^k\, \mathbf{e}_j,\, \mathbf{e}_i\rangle \\
&= \langle \sum_{k=0}^{\infty} \frac{g_\pm^{(k)}(0)}{k!} A^k\, \mathbf{e}_j,\, \mathbf{e}_i\rangle,
\end{aligned}$$

which completes the proof.

Hence, (4.4.15) is a direct consequence of the above lemma.

4.4.2 Finite Dimensional Approximation

We now construct a finite dimensional approximation to (4.4.11)–(4.4.14) which can be decoupled by simultaneous diagonalization. We note that while the infinite matrices A and $S^\pm$ commute (cf. (4.4.15)), a direct truncation does not preserve the commutativity (cf. Remark 4.4.1) which is essential for the decoupling.

Define a finite dimensional subspace of $L^2(\mathbb{R})$:

$$X_M = \text{span}\{\psi_0, \psi_1, \ldots, \psi_M\}.$$

Numerically, we shall seek the solution in the finite dimensional subspace X_M. Assume the numerical solution has the expansion

$$w_M^\pm(x, y) = \sum_{m=0}^{M} w_m^\pm(y)\psi_m(x).$$

Define a diagonal matrix $D_M = \mathrm{diag}((-\mathrm{i})^0, (-\mathrm{i})^1, (-\mathrm{i})^2, \ldots, (-\mathrm{i})^M)$ and vectors

$$\begin{aligned}
\mathbf{w}_M^{\pm}(y) &= D_M \cdot (w_0^{\pm}(y), w_1^{\pm}(y), \ldots, w_M^{\pm}(y))^{\top}, \\
\mathbf{v}_M^{\pm}(y) &= D_M \cdot (v_0^{\pm}(y), v_1^{\pm}(y), \ldots, v_M^{\pm}(y))^{\top}, \\
\boldsymbol{\rho}_M^{\pm} &= D_N \cdot (\rho_0^{\pm}, \rho_1^{\pm}, \ldots, \rho_M^{\pm})^{\top}, \\
\mathbf{h}_M &= D_M \cdot (h_1, h_2, \ldots, h_M)^{\top}.
\end{aligned}$$

Let us denote $A_M = (a_{ij})_{0 \le i,j \le M}$, where $\{a_{ij}\}$ are the entries for A, and define

$$S_M^{\pm} = \sum_{k=0}^{\infty} \frac{g_{\pm}^{(k)}(0)}{k!} (A_M)^k. \tag{4.4.20}$$

Remark 4.4.1 We observe that the matrices $S_M^{\pm}$ are not simply the truncation of the matrices $S^{\pm}$ given by

$$(S^{\pm})_M = \sum_{k=0}^{\infty} \frac{g_{\pm}^{(k)}(0)}{k!} (A^k)_M.$$

However, the difference between $S_M^{\pm}$ and $(S^{\pm})_M$ are high order terms which converge to zero as M increases.

Then, our finite dimensional approximation to (4.4.11)–(4.4.14) is as follows:

$$\frac{\mathrm{d}^2 \mathbf{w}_M^{\pm}(y)}{\mathrm{d}y^2} + (\kappa_{\pm}^2 I_M - A_M)\mathbf{w}_M^{\pm}(y) = \mathbf{v}_M^{\pm}(y), \tag{4.4.21}$$

with the boundary conditions

$$\frac{\mathrm{d}\mathbf{w}_M^{\pm}(y_{\pm})}{\mathrm{d}y} - S_M^{\pm}\mathbf{w}_M^{\pm}(y_{\pm}) = \boldsymbol{\rho}_M^{\pm}, \tag{4.4.22}$$

and the continuity conditions

$$\mathbf{w}_M^{+}(0) - \mathbf{w}_M^{-}(0) = 0, \tag{4.4.23}$$

and

$$\frac{\mathrm{d}\mathbf{w}_M^{+}(0)}{\mathrm{d}y} - \frac{\mathrm{d}\mathbf{w}_M^{-}(0)}{\mathrm{d}y} = \mathbf{h}_M. \tag{4.4.24}$$

Note that the above system of $(M+1)$ equations is *coupled* together by the matrices A_M and $S_M^{\pm}$. In order to decouple the above system, we need to show the following result:

Theorem 4.4.1 The matrices A_M and $S_M^{\pm}$ are simultaneously diagonalizable. Moreover, let $\{\lambda_j\}_{0\le j\le M}$ be the set of eigenvalues of A_M, then the eigenvalues of $S_M^{\pm}$ are given by

$$\sigma_j^{\pm} = \mathrm{i}\sqrt{\kappa_{\pm}^2 - \lambda_j}, \quad 0 \le j \le M. \tag{4.4.25}$$

Proof By definition, A_M is a real symmetric matrix. Hence, there exists an $M \times M$ orthonormal matrix Q_M such that

$$Q_M^{\top} A_M Q_M = \Lambda_M,$$

where Λ_M is an $M \times M$ diagonal matrix.

It follows from the definition of $S_M^{\pm}$ that we have

$$Q_M^{\top} S_M^{\pm} Q_M = \sum_{k=0}^{\infty} \frac{g_{\pm}^{(k)}(0)}{k!} Q_M^{\top} (A_M)^k Q_M = \sum_{k=0}^{\infty} \frac{g_{\pm}^{(k)}(0)}{k!} (Q_M^{\top} A_M Q_M)^k \tag{4.4.26}$$

$$= \sum_{k=0}^{\infty} \frac{g_{\pm}^{(k)}(0)}{k!} \Lambda_M^k := \Sigma_M^{\pm} = \mathrm{diag}\{\sigma_0^{\pm}, \sigma_1^{\pm}, \cdots, \sigma_M^{\pm}\}., \tag{4.4.27}$$

Moreover, (4.4.25) is a direct consequence of the above and (4.4.19).

By Theorem 4.4.1, the matrices $S_M^{\pm}$ and A_M can be simultaneously diagonalized by the same orthogonal matrix Q_M, i.e., there exist an orthonormal matrix Q_M and two $M \times M$ diagonal matrices Λ_M and Σ_M such that

$$Q_M^{\top} A_M Q_M = \Lambda_M \quad \text{and} \quad Q_M^{\top} S_M^{\pm} Q_M = \Sigma_M^{\pm}.$$

Denote $\tilde{\mathbf{w}}_M^{\pm}(y) = Q_M^{\top}\mathbf{w}_M^{\pm}(y), \tilde{\mathbf{v}}_M^{\pm}(y) = Q_M^{\top}\mathbf{v}_M^{\pm}(y), \tilde{\boldsymbol{\rho}}_M^{\pm} = Q_M^{\top}\boldsymbol{\rho}_M^{\pm}, \tilde{\mathbf{h}}_M = Q_M^{\top}\mathbf{h}_M$. Multiplying $Q_M^{\top}$ on both sides of (4.4.21) and using the simultaneous diagonalization property, we deduce a fully decoupled system of $M+1$ equations:

$$\frac{\mathrm{d}^2\tilde{\mathbf{w}}_M^{\pm}(y)}{\mathrm{d}y^2} + (\kappa_{\pm}^2 I_M - \Lambda_M)\tilde{\mathbf{w}}_M^{\pm}(y) = \tilde{\mathbf{v}}_M^{\pm}(y), \tag{4.4.28}$$

with the boundary conditions

$$\frac{\mathrm{d}\tilde{\mathbf{w}}_M^{\pm}(y_{\pm})}{\mathrm{d}y} - \Sigma_M^{\pm}\tilde{\mathbf{w}}_M^{\pm}(y_{\pm}) = \tilde{\boldsymbol{\rho}}_M^{\pm}, \tag{4.4.29}$$

and the continuity conditions

$$\tilde{\mathbf{w}}_M^+(0) - \tilde{\mathbf{w}}_M^-(0) = 0, \tag{4.4.30}$$

and

$$\frac{\mathrm{d}\tilde{\mathbf{w}}_M^+(0)}{\mathrm{d}y} - \frac{\mathrm{d}\tilde{\mathbf{w}}_M^-(0)}{\mathrm{d}y} = \tilde{\mathbf{h}}_M. \tag{4.4.31}$$

Once $\tilde{\mathbf{w}}_M^{\pm}$ is obtained by solving the above decoupled two-point boundary value problem (4.4.28)–(4.4.31), we can compute $\mathbf{w}_M^{\pm} = Q_M \tilde{\mathbf{w}}_M^{\pm}$.

Remark 4.4.2 We observe that, in practice, it is not required to generate the matrix $S_M^{\pm}$ explicitly, since all we need is the diagonal matrix $\Sigma_M^{\pm}$ whose elements can be computed by using (4.4.25).

4.4.3 Legendre-Galerkin Approximation

The problem (4.4.28)–(4.4.31) consists of a sequence of decoupled two-point boundary value problem which can be solved, for example, by the Legendre-Galerkin method [72]. In this section, we briefly discuss the Legendre-Galerkin method to solve the following two-point boundary value model problem, and refer to the book by Shen et al. [73] for more detail.

Consider the second-order ordinary differential equation

$$\frac{\mathrm{d}^2 u^{\pm}(y)}{\mathrm{d}y^2} + \eta^{\pm} u^{\pm}(y) = v^{\pm}(y), \tag{4.4.32}$$

together with the boundary conditions on $y = y^{\pm}$

$$\frac{\mathrm{d}u^{\pm}(y^{\pm})}{\mathrm{d}y} - \sigma^{\pm} u^{\pm}(y^{\pm}) = \rho^{\pm}, \tag{4.4.33}$$

and the continuity conditions

$$u^+(0) - u^-(0) = 0, \tag{4.4.34}$$

$$\frac{\mathrm{d}u^+(0)}{\mathrm{d}y} - \frac{\mathrm{d}u^-(0)}{\mathrm{d}y} = h. \tag{4.4.35}$$

The above one-dimensional transmission problem (4.4.32)–(4.4.35) is exactly the same as the one studied in [70], and can be efficiently solved by using a Legendre-Galerkin method (cf. [72]) described in detail in Section 4 of [70]. With a suitable choice of basis functions, the Legendre-Galerkin method leads to a sequence of sparse linear system which can be inverted in $O(N)$ operations, where N is the number of unknowns in the Legendre expansion.

4.4.4 The Complete Algorithm

We now summarize the complete algorithm and its computational complexity.

Given problem parameters: wave numbers (κ_+, κ_-), perturbation width ε, we choose the numerical parameters: M to be the number of Hermite expansion in the horizontal x direction, N to be the number of Legendre expansion in the vertical y direction, and K to be the number of Taylor expansion retained in the perturbation expansion. Then, our numerical solution can be written in the following form:

$$u^{\pm}(x,y) = \sum_{m=0}^{M}\sum_{n=0}^{N}\sum_{k=0}^{K} w^{\pm}_{m,n,k}\,\psi_m(x)\,\phi^{\pm}_n(y)\,\varepsilon^k. \tag{4.4.36}$$

Therefore, the numerical algorithm is to compute the coefficients set $\{w^{\pm}_{m,n,k}\}$ for $m = 0, ..., M, n = 0, ..., N$, and $k = 0, ..., K$, which can be summarized as follows:

Pre-computation: (independent of wavenumbers $\kappa_\pm$)

1. Compute the Hermit Gauss points $\{x_n\}_{n=0,...,M}$, Legendre-Gaussian-Lobatto collocation points $\{y^+_n\}_{n=0,...,N}$ on interval $[0, y^+]$ and $\{y^-_n\}_{n=0,...,N}$ on interval $[y^-, 0]$; ($O(N) + O(M)$ flops)

2. Compute the matrix A_M, and its eigenpair (Q_M, Λ_M). ($O(M^2)$ flops)

Then, for each incident wave:

1. Compute $\Sigma^{\pm}_M$ through (4.4.25);

2. **for** $k = 1 : K$ **do**

   for $m = 1 : M$ **do**

   Solve each one dimensional problem to obtain $\{w^{\pm}_{m,n,k}\}$ for $n = 0, ..., N$

   end for

   end for

3. Calculate $u^{\pm}(x, y)$ through (4.4.36).

The computational complexity for each k in Step 2 is of order $O(M^2N)$ which comes from the matrix-matrix multiplications involving the eigenmatrix Q_M. Hence, the total computational complexity is of order $O(M^2NK)$.

4.5 Numerical Experiments

In this section, we shall present some numerical experiments to demonstrate the efficiency and accuracy of the proposed method. Two cases are considered; one is a plane surface scattering, where the analytic solution is available and can be used for accuracy test of the numerical solution, and another is rough surface scattering.

4.5.1 Plane Surface Scattering

In [74], the fundamental solution is introduced for the two-dimensional Helmholtz equation in a two-layered medium. For the observation point $\mathbf{x} = (x_1, x_2)$ and source point $\mathbf{y} = (y_1, y_2)$, the fundamental solution of Helmholtz equation in a two-layered background medium in $\mathbb{R}^2$ satisfies

$$\Delta G(\mathbf{x}, \mathbf{y}) + \kappa^2(\mathbf{x})G(\mathbf{x}, \mathbf{y}) = -\delta(\mathbf{x} - \mathbf{y}),$$

with continuity conditions on the interface

$$\begin{aligned} G(\mathbf{x}, \mathbf{y})|_{x_2=0^+} &= G(\mathbf{x}, \mathbf{y})|_{x_2=0^-} \\ \partial_{x_2} G(\mathbf{x}, \mathbf{y})|_{x_2=0^+} &= \partial_{x_2} G(\mathbf{x}, \mathbf{y})|_{x_2=0^-} \end{aligned},$$

where the wavenumber

$$\kappa(\mathbf{x}) = \begin{cases} \kappa_1 & \text{for } x_2 > 0, \\ \kappa_2 & \text{for } x_2 < 0. \end{cases}$$

Denote $\beta_i^2 = \kappa_i^2 - \xi^2$ with $\mathrm{Im}\beta_i \geq 0$. It follows from the Fourier transform that the fundamental solution is given by

$$G(\mathbf{x}, \mathbf{y}) = \begin{cases} \Psi^{(1)}(\mathbf{x}, \mathbf{y}) + \Phi_1(\mathbf{x}, \mathbf{y}) & \text{for } x_2 > 0,\ y_2 > 0, \\ \Psi^{(2)}(\mathbf{x}, \mathbf{y}) + \Phi_2(\mathbf{x}, \mathbf{y}) & \text{for } x_2 < 0,\ y_2 < 0, \\ \Psi^{(3)}(\mathbf{x}, \mathbf{y}) & \text{for } x_2 > 0,\ y_2 < 0, \\ \Psi^{(4)}(\mathbf{x}, \mathbf{y}) & \text{for } x_2 < 0,\ y_2 > 0, \end{cases}$$

where

$$\begin{aligned} \Psi^{(1)}(\mathbf{x}, \mathbf{y}) &= \frac{\mathrm{i}}{4\pi} \int_{-\infty}^{\infty} \frac{1}{\beta_1} \frac{\beta_1 - \beta_2}{\beta_1 + \beta_2} e^{\mathrm{i}\beta_1(x_2+y_2)} e^{\mathrm{i}\xi(x_1-y_1)} d\xi, \\ \Psi^{(2)}(\mathbf{x}, \mathbf{y}) &= \frac{\mathrm{i}}{4\pi} \int_{-\infty}^{\infty} \frac{1}{\beta_2} \frac{\beta_2 - \beta_1}{\beta_1 + \beta_2} e^{-\mathrm{i}\beta_2(x_2+y_2)} e^{\mathrm{i}\xi(x_1-y_1)} d\xi, \\ \Psi^{(3)}(\mathbf{x}, \mathbf{y}) &= \frac{\mathrm{i}}{2\pi} \int_{-\infty}^{\infty} \frac{e^{\mathrm{i}(\beta_1 x_2 - \beta_2 y_2)}}{\beta_1 + \beta_2} e^{\mathrm{i}\xi(x_1-y_1)} d\xi, \\ \Psi^{(4)}(\mathbf{x}, \mathbf{y}) &= \frac{\mathrm{i}}{2\pi} \int_{-\infty}^{\infty} \frac{e^{\mathrm{i}(\beta_1 y_2 - \beta_2 x_2)}}{\beta_1 + \beta_2} e^{\mathrm{i}\xi(x_1-y_1)} d\xi, \end{aligned}$$

and Φ_i is the fundamental solution of the Helmholtz equation in homogeneous background medium in $\mathbb{R}^2$ with wavenumber κ_i, i.e.,

$$\Phi_i(\mathbf{x}, \mathbf{y}) = \frac{\mathrm{i}}{4} H_0^{(1)}(\kappa_i |\mathbf{x} - \mathbf{y}|), \quad i = 1, 2.$$

Here $H_0^{(1)}$ is the Hankel function of the first kind with order zero.

We consider the case where the surface is a plane, i.e., $f(x) = 0$. We can compare numerical solution with the analytic solution given above. Recall that the point $(x_0; y_0)$ is where the source is placed. In this section, we always assume the point source is placed at $(x_0;\ y_0) = (0.0; 1.5)$ and the transparent boundaries are put at $y^+ = 1$ and $y^- = -1$.

First, we investigate the convergence of the series solution in the horizontal x direction. We fix $N = 40$ and vary M with four different wavenumber cases:

$$\begin{aligned}
&\text{Case1} : (\kappa_+,\, \kappa_-) = (10.5 + 1.0\text{i},\, 20.5 + 1.0\text{i}),\\
&\text{Case2} : (\kappa_+,\, \kappa_-) = (1.5 + 1.0\text{i},\, 2.5 + 1.0\text{i}),\\
&\text{Case3} : (\kappa_+,\, \kappa_-) = (10.5 + 0.5\text{i},\, 20.5 + 0.5\text{i}),\\
&\text{Case4} : (\kappa_+,\, \kappa_-) = (1.5 + 0.5\text{i},\, 2.5 + 0.5\text{i}).
\end{aligned}$$

The results are shown in Figure 4.2 (left), which plots the $L^2(\Omega)$ error of the numerical solution again the number of truncation in the x-direction M. From the results, we can observe that the numerical solutions converge exponentially as M is increased. We point out that the wavenumbers with large real and small imaginary values require much larger M in order to maintain the same order of accuracy. However, for those wavenumbers with either big real and imaginary values or small real and imaginary values, e.g., Case 1 and Case 4, the impact of the imaginary value of the wavenumber will eventually affect the numerical accuracy.

Next, we investigate the convergence of the series solution in vertical y-direction. In this test, we only vary the parameter N, and take a sufficiently large M, e.g., $M = 160$, such that the approximation error is negligibly small in the x-direction. We consider the same four cases as the previous investigation. The results are shown in Figure 4.2 (right), which plot the $L^2(\Omega)$ error of the numerical solution against the number of truncation in the y-direction N. Again, we notice an exponential convergence as N is increased, and that the wavenumbers with large real and small imaginary values require much larger N to reach the same level of accuracy. Similarly, for the wavenumbers with either big real and imaginary values or small real and imaginary values, e.g., Case 1 and Case 4, the impact of the imaginary value of the wavenumber will eventually affect the numerical accuracy.

Finally, we investigate the convergence of the series solution with respect to the wavenumber. We fix the real part of $\kappa_\pm$, e.g., $\text{Re}(\kappa_+) = 1.5$ and $\text{Re}(\kappa_-) = 2.5$, and vary both of the imaginary parts of $\kappa_\pm$ from 0.1 to 1. The results are displayed in

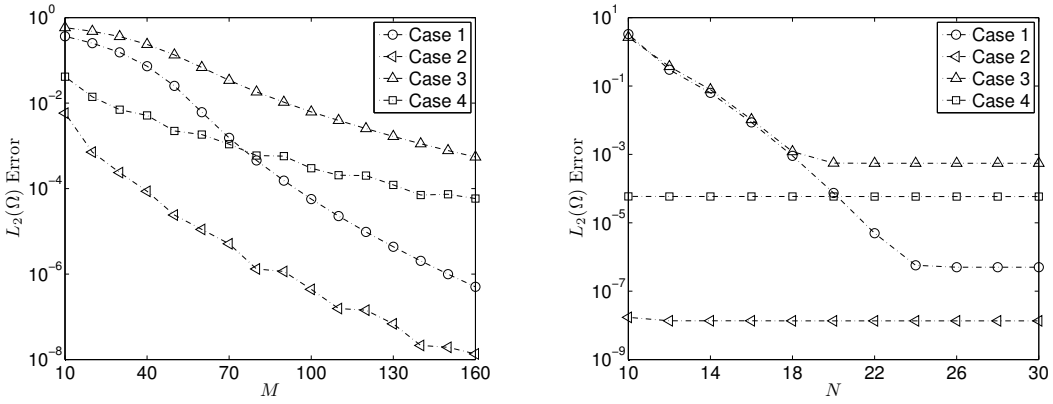


Figure 4.2. The $L^2(\Omega)$ error of the numerical solution is plotted against the number of truncation terms for flat surface scattering. (left) The error is plotted against the truncation term in the horizontal x-direction M; (right) The error is plotted against the truncation term in the vertical y-direction N.

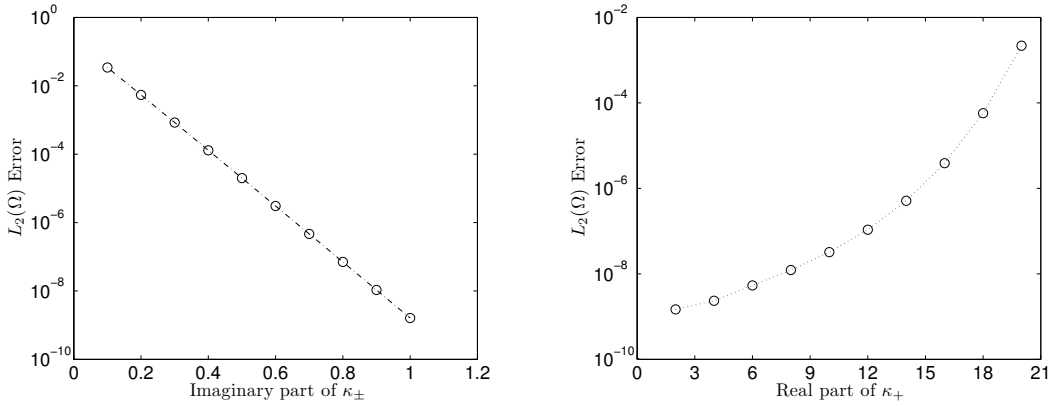


Figure 4.3. The $L^2(\Omega)$ error of the numerical solution is plotted against the wavenumbers for flat surface scattering. (left) The error is plotted against the real parts of the wavenumbers; (right) The error is plotted against the imaginary parts of the wavenumbers.

Figure 4.3 (left) for fixed $M = 200$ and $N = 80$. Then we fix the imaginary part of $\kappa_{\pm}$, e.g., $\mathrm{Im}(\kappa_+) = 1$ and $\mathrm{Im}(\kappa_-) = 1$, and vary both of the real parts of $\kappa_{\pm}$ from $\mathrm{Re}(\kappa_+) = 1$ to 20 and $\mathrm{Re}(\kappa_-) = 2\mathrm{Re}(\kappa_+)$. The result are displayed in Figure 4.3 (right) with fixed truncation terms $M = 200$ and $N = 80$. As expected, the error decreases as the imaginary part of the wavenumber increases, while the error increases as the real part of the wavenumber increases.

4.5.2 Rough Surface Scattering

In this subsection, we will investigate the case of rough surface scattering and determine how the numerical accuracy depends on the parameter K, i.e., the term in the series solution. We will fix the parameters M and N such that the approximation error is negligibly small in terms of M and N. To test the convergence of the method, we denote a relative $L^2(\Omega)$ error

$$\mathrm{E}_K = \frac{\|u_K - u_{K-1}\|_{L^2(\Omega)}}{\|u_K\|_{L^2(\Omega)}}.$$

First, we consider the wavenumber data set of Case 2, i.e., $(\kappa^+, \kappa^-) = (1.5 + 1.0\text{i}, 2.5 + 1.0\text{i})$ and choose the function $g_1(x) = \cos(x)$ to represent the rough surface. We fix $\varepsilon = 0.1$, $M = 100$, $N = 30$, and vary K from 1 to 16. The convergent result is displayed in the column of Test 1 in Table 4.1 . We also change ε from 0.2 to 0.8. The convergent results with respect to different ε are displayed in Figure 4.4, which plots the relative $L^2(\Omega)$ error E_K against the number of truncation in the series solution K. From the results in Figure 4.4, we can observe that the convergence rate highly depends on the value of ε when fixing all the other parameters, and smaller ε leads to faster convergence with a few iterations.

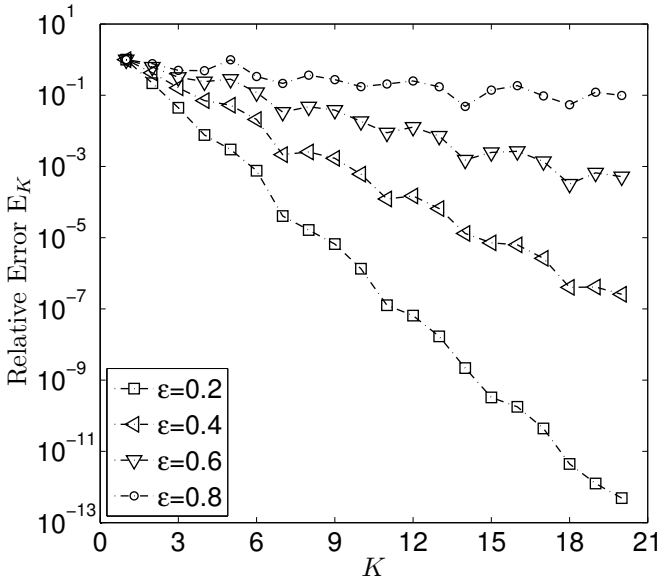


Figure 4.4. Relative $L^2(\Omega)$ error is plotted against the number of truncation K in the series solution.

Second, we still consider the same wavenumber data set of Case 2, i.e., $(\kappa^+, \kappa^-) = (1.5 + 1.0\text{i}, 2.5 + 1.0\text{i})$, choose the same truncation terms $M = 100$ and $N = 30$, and vary K from 1 to 16. We choose the function $g_2(x) = \cos(4x) + 2\cos(2x) + 4\cos(x)$ for the rough surface with two different values of parameter ε: $\varepsilon_1 = 0.1$ and $\varepsilon_2 = 0.1/7$ such that $\max\{\varepsilon_2 g_2(x)\} = 0.1$. The convergent results are displayed in the column

Test 2 and Test 3 in Table 4.1. The convergence for Test 2, corresponding to larger perturbation parameter $\varepsilon_1 = 0.1$, is much slower than the convergence for Test 3, corresponding to smaller perturbation parameter $\varepsilon_2 = 0.1/7$.

Next, we consider the wavenumber data set of Case 1, i.e., $(\kappa^+, \kappa^-) = (10.5 + 1.0\text{i}, 20.5 + 1.0\text{i})$ for the same $M = 160, N = 30$, and vary K from 1 to 16. The rough surface is chosen as $g_1(x) = \cos(x)$ and $\varepsilon = 0.1$. The convergent results are displayed in the column Test 4 in Table 4.1.

From those results displayed in Table 4.1, we can observe that for Test 1 and Test 2, although we consider the same wave numbers (κ^+, κ^-), parameters (M, N) and same value of ε but with different profiles $g_1(x)$ and $g_2(x)$, the convergence rate of Test 1 is apparently much faster than that of Test 2. The reason is because the real profile of the surface is $f(x) = \varepsilon g(x)$, but the $\max|g_2(x)|$ is almost seven times of $\max|g_1(x)|$, so when we decrease the value of ε to $0.1/7$ in Test 3, we can see that the convergence rate of Test 1 and Test 3 are almost the same as we increase K. Therefore the maximum height of the perturbed profile plays more important role on the convergence of our numerical algorithm than its shape, which is also consistent with the results in Figure 4.4. Comparing the results of Test 1 and Test 4, we can observe that Test 4 with large wave numbers converges much slower than Test 1 with small wave numbers. So we can conclude that the value of the wavenumber is also another important factor on the convergence rate.

Finally, we consider three different examples given in Table 4.2. The contour plot of the total field are displayed in Figure 4.5, Figure 4.6, and Figure 5.4, respectively.

4.6 Conclusion

We constructed and implemented a new spectral method for solving the two-dimensional acoustic wave scattering problem by unbounded rough surfaces. The main difficulty of the problem is that the non-local boundary conditions prevent us from decoupling the two-dimensional system to a sequence of one-dimensional prob-

Table 4.1
Convergence test for different wavenumbers and perturbation parameter ε for rough surface scattering.

K	Test 1	Test 2	Test 3	Test 4
1	1.0E+00	1.0E+00	1.0E+00	1.0E+00
2	1.09E-01	5.71E-01	8.68E-02	6.07E-01
3	1.14E-02	3.03E-01	4.03E-03	3.83E-01
4	8.67E-04	3.56E-01	9.33E-04	3.71E-01
5	1.75E-04	3.74E-01	1.33E-04	1.27E-01
6	2.47E-05	9.12E-02	1.00E-05	1.87E-01
7	7.31E-07	8.15E-02	3.73E-07	9.91E-02
8	1.14E-07	1.01E-01	1.15E-07	9.27E-02
9	2.51E-08	3.29E-02	1.00E-08	6.64E-02
10	2.69E-09	3.64E-02	6.76E-10	3.22E-02
11	1.41E-10	5.20E-02	1.28E-10	3.89E-02
12	2.91E-11	2.80E-02	1.58E-11	1.57E-02
13	4.11E-12	1.03E-02	8.80E-13	1.99E-02
14	2.93E-13	2.18E-02	1.51E-13	1.25E-02
15	1.70E-14	1.56E-02	2.40E-14	8.52E-03
16	5.01E-15	4.49E-03	2.00E-15	8.26E-03

Table 4.2
Three test examples for rough surface scattering.

Test	(κ^+, κ^-)	ε	$g(x)$	(M, N, K)
Case 5	$(1.5 + 1.0\mathrm{i}, 2.5 + 1.0\mathrm{i})$	0.1	$g_1(x)$	$(100, 30, 20)$
Case 6	$(5.5 + 1.0\mathrm{i}, 10.5 + 1.0\mathrm{i})$	0.2	$g_1(x)$	$(160, 30, 20)$
Case 7	$(1.5 + 0.5\mathrm{i}, 2.5 + 1.0\mathrm{i})$	0.1	$g_2(x)$	$(150, 40, 20)$

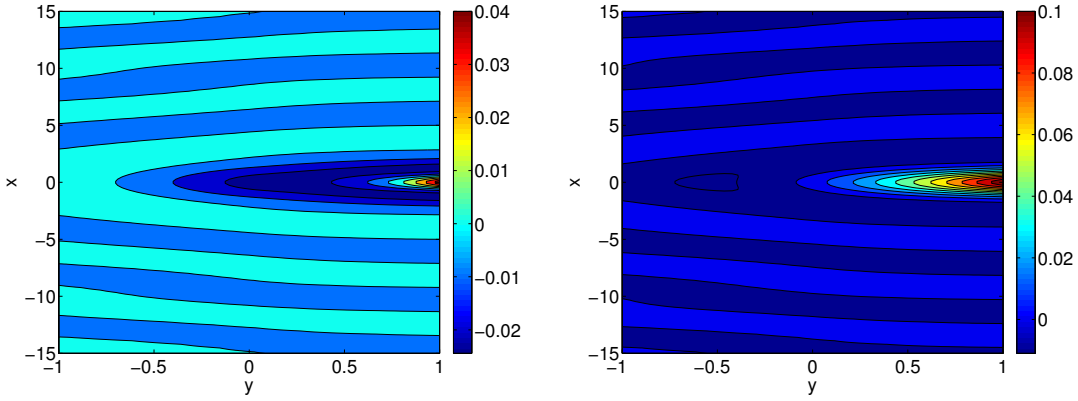


Figure 4.5. Contour plot of the total field for Case 5. (left) real part of the total field; (right) imaginary part of the total field.

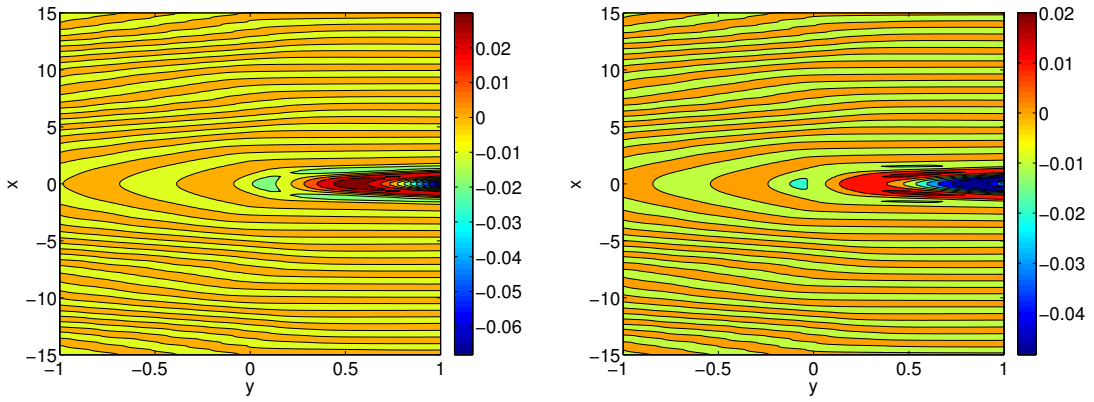


Figure 4.6. Contour plot of the total field for Case 6. (left) real part of the total field; (right) imaginary part of the total field.

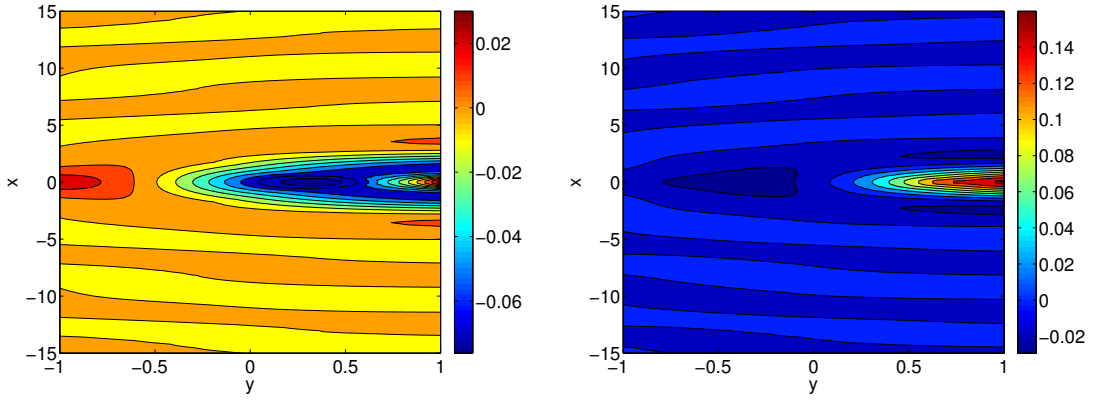


Figure 4.7. Contour plot of the total field for Case 7. (left) real part of the total field; (right) imaginary part of the total field.

lems with a usual approach. The main novelty of the proposed method is to expand the solution using the Hermite orthonormal basis functions in the Fourier space, and to simultaneously diagonalize the two coupling matrices by using the essential property that the Hermite functions are eigenfunctions of the Fourier transform. The combined approach allows us to reduce the original two-dimensional boundary value problem into a sequence of fully decoupled the one-dimensional Helmholtz equations, with piecewise constant wavenumbers, that can be efficiently solved by using a Legendre-Galerkin method.

We investigated the errors of the numerical solution in terms of the horizontal truncation term M, vertical truncation term N, power series truncation term K, and the wavenumbers $\kappa_{\pm}$, for both the plane surface scattering and the rough surface scattering. The numerical results indicate that the method is efficient, accurate, and well suited for the unbounded rough surface scattering problem.

It is clear to see that the current approach can be extended to handle the two-dimensional multi-layered unbounded rough surface scattering. We plan to extend the method to the electromagnetic wave scattering by unbounded rough surfaces, where the three-dimensional Maxwell equations have to be considered.

CHAPTER 5. UNCONDITIONAL STABLE PRESSURE-CORRECTION SCHEMES FOR NON-LINEAR NO-SLIP FLUID-STRUCTURE PROBLEM

In this chapter, we consider the numerical approximation of the nonlinear fluid-structure interaction (FSI). Our goal is to construct unconditionally energy stable schemes that are also computationally very efficient. This is achieved by using pressure-correction approaches with proper pressure boundary conditions at the interface. More precisely, we construct unconditionally stable standard and rotational pressure-correction schemes for the FSI problem with a fixed interface. In addition, these schemes are computationally very efficient, as they lead to, at each time step, a coupled linear elliptic system for the velocity and displacement in the whole region and a discrete Poisson equation in the fluid region. We validate these schemes by using a Fourier-Legendre spatial discretization for the FSI problem in a periodic channel.

5.1 Introduction

Fluid-Structure Interaction (FSI) plays an important role in many scientific/engineering applications, e.g., design of engineering systems, blood flow in human arteries, etc. It has been extensively studied in recent years both analytically and computationally (cf. [26–28, 75] and the references therein).

There are two main approaches, *monolithic* and *partitioned* methods, for solving FSI problems numerically. The partitioned approach (cf., for instance, [30–32, 76]) solves the fluid and structure dynamics separately with explicit interface conditions. While each subproblem can be solved efficiently by existing algorithms, the explicit treatment of the interface condition may lead to instability and very restrictive time

step constraint. In contrast, the monolithic approach (cf., for instance, [33–35]) simultaneously solves the fluid and structure dynamics coupled by the implicit interface conditions. This type of schemes usually have good stability properties, but at each time step, a nonlinear coupled system has to be solved, and due to the presence of the pressure in the coupled system, it is usually difficult to design effective iterative scheme to solve the nonlinear coupled system. Some authors have also developed reduced monolithic methods which are based on semi-implicit coupling at the interface (cf., for instance, [77]).

For fluid problems, an effective approach to decouple the computation of the pressure from that of the velocity is to use a so called projection type method, originally proposed by Chorin and Temam in the late 60's. A comprehensive review on various projection type methods can be found in [36]. Naturally, many authors have considered to employ a projection type method for the FSI problem (cf., for instance, [37, 78]). However, a main difficulty in the design of a projection method is what boundary condition to use for the pressure at the interface. It is well known that a proper boundary condition, at the Dirichlet part of the boundary, for the pressure Poisson equation in a projection type method is the homogeneous Neumann boundary condition. Indeed, most existing projection type scheme for FSI problem also use, explicitly or implicitly, Neumann type boundary condition for the pressure Poisson equation at the interface. However, we are not aware of any rigorous proof of unconditional stability for any projection type scheme applied to the FSI problem.

In [79], the authors proposed and analyzed pressure-correction projection schemes for Navier-Stokes equations with open boundary where the usual stress-free boundary condition is applied. It is shown that the proper boundary condition at the open boundary is of Dirichlet type instead of Neumann type. Two schemes are constructed in [79], one is based on the standard pressure-correction which leads to poor accuracy at the open boundary, the other is based on the rotational pressure-correction and with a proper Dirichlet boundary condition at the open boundary. It is shown in [79] that both the standard and rotational pressure-correction projection schemes, when

applied to the time-dependent Stokes problem, are unconditionally stable, but the rotational version leads to much better accuracy. Since one of matching interface condition for the FSI problem is related to the stress, it makes sense to extend the approach in [79] for problems with open boundary to the FSI problems.

Besides the difficulty associated with the pressure boundary condition on the interface, another major difficulty is to prove the unconditional stability of the rotational pressure-correction scheme for the nonlinear FSI problem. The original stability proof of the rotational pressure-correction scheme in [80] was only valid for Stokes problems. An essential step of the proof was to take the "discrete time derivative" of the scheme. Unfortunately, this proof can not be extended to the nonlinear case. In [81], the authors constructed an unconditionally stable rotational velocity-correction scheme for the Navier-Stokes equations. However, they only provided a stability proof for the linear Stokes equations, while showed numerically that the scheme was unconditionally stable. In [82], the author proposed a Gauge-Uzawa approach for the rotational pressure-correction scheme of the Navier-Stokes equations, and proved that the scheme is unconditionally stable. We shall extend the approach in [82] for the Gauge-Uzawa scheme of the Navier-Stokes equations to the rotational pressure-correction schemes for the FSI problem.

We consider in this paper a simple model of the FSI problem where the movement of the interface is assumed infinitesimal so the interface is treated as fixed. This nonlinear FSI problem captures many of the essential difficulties of the more general FSI problems with moving interface, and its well-posedness has been studied in [29]. In [83], the authors studied a semi-discrete (in space) finite-element method for a linear FSI problem with a fixed interface. We shall be mainly concerned with semi-discrete (in time) projection type schemes for the nonlinear FSI problem with a fixed interface.

Based on the above considerations, we construct in this paper several monolithic schemes based on a pressure-correction approach for the FSI model with fixed interface. Our schemes will be computationally very efficient. More precisely, in the first

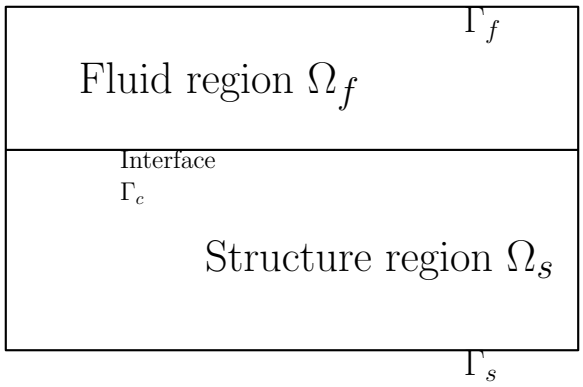


Figure 5.1. Geometry discription for fluid-structure problem

step of our schemes, we solve a coupled, but elliptic, system for an intermediate fluid velocity and the structure displacement, then in the second step, we solve a Poisson equation for the fluid pressure and obtain the fluid velocity with a simple correction. We shall also prove rigorously that these schemes are unconditionally stable.

The rest of the paper is organized as follows. In the next section, we describe the governing equations for our FSI model, formulate its weak form and the energy dissipation law. In Section 3, we construct standard and rotational pressure-correction scheme for the FSI problem, and prove their unconditional stability. Then, in Section 4, we describe a generic approach for spatial discretization as well as a Fourier-Legendre method for a special case of a periodic channel. We present some numerical results in Section 5 to validate our numerical schemes and to demonstrate their temporal accuracy.

5.2 Governing Equations

We consider the following model for interaction of a viscous fluid with an elastic body in a two- or three-dimensional bounded domain Ω, with the fluid region Ω_f, the solid region Ω_s and the interface Γ_c so we have $\Omega = \Omega_f \cup \Omega_s \cup \Gamma_c$, we also denote $\Gamma_f = \partial\Omega_f \backslash \Gamma_c$ and $\Gamma_s = \partial\Omega_s \backslash \Gamma_c$ (cf. Fig. 5.1).

We assume that the interface undergoes infinitesimal displacements, i.e., Γ_c is fixed. The more complicated situation with moving interface will be considered in a forthcoming paper.

In the fluid region Ω_f, we have the Navier-Stokes equations:

$$\rho_f u_t - \text{ div } \epsilon(u) + (u \cdot \nabla)u + \nabla p = \rho_f f_1 \qquad in \quad \Omega_f \times (0,T) \tag{5.2.1a}$$

$$\text{div } u = 0 \qquad in \quad \Omega_f \times (0,T) \tag{5.2.1b}$$

$$u = 0 \qquad on \quad \Gamma_f \times (0,T) \tag{5.2.1c}$$

$$u|_{t=0} = u_0 \qquad in \quad \Omega_f \tag{5.2.1d}$$

where u denotes the fluid velocity, p the fluid pressure, u_0 the given initial velocity, f_1 the given body force per unit mass, $\epsilon(u) = \frac{\mu}{2}(\nabla u + \nabla u^T)$ the strain tensor, ρ_f and μ the constant fluid density and viscosity.

In the solid region Ω_s, we have the wave equation for linear elasticity:

$$\rho_s w_{tt} - \text{ div } \sigma(w) = \rho_s f_2 \qquad in \quad \Omega_s \times (0,T) \tag{5.2.2a}$$

$$w = 0 \qquad on \quad \Gamma_s \times (0,T) \tag{5.2.2b}$$

$$w(\cdot, 0) = w_0 \qquad in \quad \Omega_s \tag{5.2.2c}$$

$$w_t(\cdot, 0) = w_1 \qquad in \quad \Omega_s \tag{5.2.2d}$$

where w denotes the displacement of the solid, w_0 and w_1 the given initial data, and $\sigma(w)$ the elastic stress tensor,and f_2 the given loading force per unit mass, λ and μ_2 the Lamé constants, ρ_s the constant solid density.

Across the fixed interface Γ_c between the fluid and solid, the velocity and the stress vector are required to be continuous, i.e.,

$$w_t = u, \quad on\ \Gamma_c \times (0,T) \tag{5.2.3}$$

and

$$\sigma(w) \cdot \mathbf{n} = \epsilon(u) \cdot \mathbf{n} - p\mathbf{n} - \frac{1}{2}(u \cdot \mathbf{n})u, \quad on\ \Gamma_c \times (0,T) \tag{5.2.4}$$

where $\mathbf{n}$ denotes the outward normal vector along Γ_c w.r.t. Ω_s. For instance, if $\Gamma_c = \{(x, y)|y = 0\}$, then $\mathbf{n} = (0, 1)$.

For simplicity, we take in this paper $\rho_f = \rho_s = 1$, $f_1 = f_2 = 0$. We further take $\lambda = 1$ and $\mu_2 = 0$ which implies div $\sigma(w) = \Delta w$, and the interface condition (5.2.4) reduces to

$$\frac{\partial w}{\partial \mathbf{n}} = \mu \frac{\partial u}{\partial \mathbf{n}} - p\mathbf{n} - \frac{1}{2}(u \cdot \mathbf{n})u, \quad on\ \Gamma_c \times (0, T). \tag{5.2.5}$$

In order to derive a weak formulation for (5.2.1)- (5.2.2), we need to introduce some notations. Let us denote by $H^k(\Omega)$ and $H_0^k(\Omega)$ (for $k \geq 0$) the standard Sobolev spaces, equipped with the standard norm $\|\cdot\|_{k,\Omega}$. In particular, we denote $L^2(\Omega) = H^0(\Omega)$ with the associated norm $\|\cdot\|$. We will use $\mathbf{H^k}(\Omega_f)$ to denote the vector-valued Sobolev spaces. We also denote

$$H^1_{0,\Gamma_f}(\Omega_f) = \{v \in H^1(\Omega_f) : v|_{\Gamma_f} = 0\}, \quad H^1_{0,\Gamma_s}(\Omega_s) = \{v \in H^1(\Omega_s) : v|_{\Gamma_s} = 0\}.$$

Then, a weak solution (u, p, w) for (5.2.1)-(5.2.2) will satisfy

$$\begin{aligned}(u_t + (u \cdot \nabla)u, \varphi)_{\Omega_f} + (\mu \nabla u, \nabla \varphi)_{\Omega_f} - (p, \text{div}\varphi)_{\Omega_f} \qquad & (5.2.6a)\\ +(\mu \frac{\partial u}{\partial \mathbf{n}} - p \cdot \mathbf{n}, \varphi)_{\Gamma_c} = 0, \quad \forall \varphi \in \mathbf{H}^1_{0,\Gamma_f}(\Omega_f),& \\ (\text{div}u, q)_{\Omega_f} = 0, \quad \forall q \in L^2(\Omega_f), \qquad & (5.2.6b)\\ (w_{tt}, \psi)_{\Omega_s} + (\nabla w, \nabla \psi)_{\Omega_s} - (\frac{\partial w}{\partial \mathbf{n}}, \psi)_{\Gamma_c} = 0, \quad \forall \psi \in \mathbf{H}^1_{0,\Gamma_s}(\Omega_s) \qquad & (5.2.6c)\end{aligned}$$

with the interface conditions (5.2.3) and (5.2.5) on Γ_c .

We can reformulate the above, using (5.2.5), to:

$$\begin{aligned}(u_t + (u \cdot \nabla)u, \varphi)_{\Omega_f} + (\mu \nabla u, \nabla \varphi)_{\Omega_f} - (p, \text{div}\varphi)_{\Omega_f} \qquad & (5.2.7a)\\ +(\frac{\partial w}{\partial \mathbf{n}} + \frac{1}{2}(u \cdot \mathbf{n})u, \varphi)_{\Gamma_c} = 0, \quad \forall \varphi \in \mathbf{H}^1_{0,\Gamma_f}(\Omega_f),& \\ (\text{div}u, q)_{\Omega_f} = 0, \quad \forall q \in L^2_0(\Omega_f), \qquad & (5.2.7b)\\ (w_{tt}, \psi)_{\Omega_s} + (\nabla w, \nabla \psi)_{\Omega_s} - (\frac{\partial w}{\partial n}, \psi)_{\Gamma_c} = 0, \quad \forall \psi \in \mathbf{H}^1_{0,\Gamma_s}(\Omega_s) \qquad & (5.2.7c)\end{aligned}$$

with $u = w_t$ on the interface Γ_c.

Setting $\varphi = u, \psi = w_t$ in (5.2.7a) and (5.2.7c), using the identity (note that $\mathbf{n}$ below is the inward normal along Γ_c w.r.t. Ω_f)

$$((u \cdot \nabla)v, v)_{\Omega_f} = -\frac{1}{2}((u \cdot \mathbf{n})v, v)_{\partial\Omega_f}, \text{ if } \mathrm{div} u = 0, \tag{5.2.8}$$

and summing up the two resultant equations, we obtain

$$\frac{1}{2}\partial_t \|u\|^2_{\Omega_f} + \mu\|\nabla u\|^2_{\Omega_f} + \frac{1}{2}\partial_t \|w_t\|^2_{\Omega_f} + \frac{1}{2}\partial_t \|\nabla w\|^2_{\Omega_s} = 0,$$

or equivalently

$$\partial_t \left\{ \|u\|^2_{\Omega_f} + \|w_t\|^2_{\Omega_f} + \|\nabla w\|^2_{\Omega_s} \right\} = -2\mu\|\nabla u\|^2_{\Omega_f} \leq 0, \tag{5.2.9}$$

where

$$E(u, w, w_t) := \|u\|^2_{\Omega_f} + \|w_t\|^2_{\Omega_f} + \|\nabla w\|^2_{\Omega_s} \tag{5.2.10}$$

is the total energy of the FSI system.

For the well posedness of the system (5.2.7), we refer to [84].

5.3 Time Discretization

For FSI problems, it is very important to design numerical schemes which have good, preferably unconditional, stability property. Usually, this is achieved by fully coupled, implicit schemes which require solving, at each time steps, a coupled, nonlinear, saddle-point system.

We construct in this section two time discretization schemes based on a pressure-correction approach. One is a first-order semi-implicit scheme using the standard pressure-correction technique, and the other is a second-order semi-implicit scheme with rotational pressure-correction. These schemes are unconditionally stable, and lead to, at each time step, a coupled, linear elliptic system in Ω and a pressure Poisson equation in Ω_f, which can be efficiently solved by standard numerical methods. The stability analysis for each scheme is carried out in this section.

5.3.1 Standard Pressure-Correction Scheme

We first construct a first-order scheme for the FSI problem based on the standard pressure-correction scheme for Navier-stokes problem with open boundary condition [79, 85]:

Step 1 : Given (u^n, p^n, w^n), compute $\widetilde{u}^{n+1} \in \mathbf{H}^1_{0,\Gamma_f}(\Omega_f)$ and $w^{n+1} \in \mathbf{H}^1_{0,\Gamma_s}(\Omega_s)$ by solving

$$\begin{aligned}&(\tfrac{\widetilde{u}^{n+1}-u^n}{\triangle t}, \varphi)_{\Omega_f} + (\mu\nabla\widetilde{u}^{n+1}, \nabla\varphi)_{\Omega_f} + \tfrac{1}{2}((u^n\cdot\mathbf{n})\widetilde{u}^{n+1}, \varphi)_{\Gamma_c} \\ &\quad +((u^n\cdot\nabla)\widetilde{u}^{n+1}, \varphi)_{\Omega_f} - (p^n, \mathrm{div}\varphi)_{\Omega_f} + (\tfrac{\partial w^{n+1}}{\partial \mathbf{n}}, \varphi)_{\Gamma_c} = 0 \quad \forall\varphi \in \mathbf{H}^1_{0,\Gamma_f}(\Omega_f),\end{aligned} \tag{5.3.1a}$$

$$\widetilde{u}^{n+1} = \frac{w^{n+1}-w^n}{\triangle t} \quad on \quad \Gamma_c \tag{5.3.1b}$$

$$(\frac{w^{n+1}-2w^n+w^{n-1}}{\triangle t^2}, \psi)_{\Omega_s} + (\nabla w^{n+1}, \nabla\psi)_{\Omega_s} - (\frac{\partial w^{n+1}}{\partial \mathbf{n}}, \psi)_{\Gamma_c} = 0 \quad \forall\psi \in \mathbf{H}^1_{0,\Gamma_s}(\Omega_s) \tag{5.3.1c}$$

This is a coupled, linear elliptic system for $(\widetilde{u}^{n+1}, w^{n+1})$, with the coupling condition at the interface Γ_c. Hence, it can be efficiently solved, for example, by a standard domain decomposition approach (cf., for instance, [86, 87]).

Step 2 : Compute $u^{n+1} \in \mathbf{H}^1(\Omega_f)$ and $p^{n+1} \in H^1(\Omega_f)$ by solving

$$\frac{u^{n+1}-\widetilde{u}^{n+1}}{\triangle t} + \nabla(p^{n+1}-p^n) = 0 \tag{5.3.2a}$$

$$\mathrm{div}\ u^{n+1} = 0, \quad in \quad \Omega_f \tag{5.3.2b}$$

$$u^{n+1}\cdot\mathbf{n}|_{\Gamma_f} = 0 \text{ and } p^{n+1}|_{\Gamma_c} = p^n|_{\Gamma_c} \tag{5.3.2c}$$

We observe that a Dirichlet boundary condition is imposed for p^{n+1} on the interface Γ_c, as opposed to the usual Neumann boundary condition in a pressure-correction formulation. This is due to the interface condition (5.2.5) which is similar to the open boundary condition considered in [79].

We denote $H^1_{0,\Gamma_c}(\Omega_f) = \{q \in H^1(\Omega_f)\backslash\mathbb{R}, q|_{\Gamma_c} = 0\}$. Then, the above system is equivalent to: Find $(p^{n+1} - p^n) \in H^1_{0,\Gamma_c}(\Omega_f)$ such that

$$(\nabla(p^{n+1} - p^n), \nabla q) = -\frac{1}{\triangle t}(\nabla \cdot \widetilde{u}^{n+1}, q), \quad \forall q \in H^1_{0,\Gamma_c}(\Omega_f), \tag{5.3.3a}$$

$$u^{n+1} = \widetilde{u}^{n+1} - \triangle t \nabla(p^{n+1} - p^n). \tag{5.3.3b}$$

Hence, we only have to solve a Poisson equation at this step.

For the above scheme, we have the following result:

Theorem 5.3.1 The scheme (5.3.1)-(5.3.3), with $p^0|_{\Gamma_c} = 0$, is unconditionally stable. More precisely, if we define the discrete energy

$$E^n = \|u^n\| + \|\delta_t w^n\|^2 + \|\nabla w^n\|^2 + (\triangle t)^2 \|\nabla p^n\|^2, \tag{5.3.4}$$

then we have, for all $n \geq 0$,

$$E^{n+1} - E^n + \|\widetilde{u}^{n+1} - u^n\|^2 + 2\mu\triangle t\|\nabla\tilde{u}^{n+1}\|^2 + \triangle t^2\|\delta^2_{tt} w^{n+1}\|^2 + \triangle t^2\|\nabla(\delta_t w^{n+1})\|^2 \leq 0.$$

Proof To simplify the notations, we define the discrete time derivatives $\delta_t u^{n+1} := \frac{u^{n+1}-u^n}{\triangle t}$ and $\delta^2_{tt} u^{n+1} := \frac{\delta_t u^{n+1} - \delta_t u^n}{\triangle t} = \frac{u^{n+1}-2u^n+u^{n-1}}{\triangle t^2}$ for any sequence $\{u^k\}$.

Taking $\varphi = 2\widetilde{u}^{n+1}$ in (5.3.1a), $\psi = 2\delta_t w^{n+1}$ in (5.3.1c), and taking the inner product of (5.3.2a) with $q = 2\triangle t \nabla p^n$, then summing up the three relations, we obtain:

$$\begin{aligned}
&\tfrac{1}{\triangle t}\{\|\widetilde{u}^{n+1}\|^2 - \|u^n\|^2 + \|\widetilde{u}^{n+1} - u^n\|^2\} + 2\|\nabla\widetilde{u}^{n+1}\|^2 - 2(p^n,\ \text{div}\ \widetilde{u}^{n+1})_{\Omega_f} \\
&+\tfrac{1}{\triangle t}\{\|\delta_t w^{n+1}\|^2 - \|\delta_t w^n\|^2 + \|\delta_t w^{n+1} - \delta_t w^n\|^2\} \\
&+\tfrac{1}{\triangle t}\{\|\nabla w^{n+1}\|^2 - \|\nabla w^n\|^2 + \triangle t^2\|\nabla\delta_t w^{n+1}\|^2\} = 0
\end{aligned} \tag{5.3.5}$$

Rewrite (5.3.2a) as

$$\frac{u^{n+1}}{\sqrt{\triangle t}} + \sqrt{\triangle t}\nabla p^{n+1} = \frac{\widetilde{u}^{n+1}}{\sqrt{\triangle t}} + \sqrt{\triangle t}\nabla p^n. \tag{5.3.6}$$

Taking inner product with itself from both sides and integrating by parts, thanks to $p^k|_{\Gamma_c} = 0$ for all k (due to $p^0|_{\Gamma_c} = 0$), and $\widetilde{u}^{n+1} \cdot \mathbf{n}|_{\Gamma_f} = 0 = u^{n+1} \cdot \mathbf{n}|_{\Gamma_f}$, we obtain

$$\frac{1}{\triangle t}\|u^{n+1}\|^2 + \triangle t\|\nabla p^{n+1}\|^2 = \frac{\|\widetilde{u}^{n+1}\|^2}{\triangle t} + \triangle t\|\nabla p^n\|^2 - 2(p^n, \text{div}\widetilde{u}^{n+1})_{\Omega_f}. \tag{5.3.7}$$

Summing up (5.3.5) and (5.3.7), we obtain

$$\begin{aligned}
&\tfrac{1}{\triangle t}\{\|u^{n+1}\|^2 - \|u^n\|^2 + \|\widetilde{u}^{n+1} - u^n\|^2\} + 2\|\nabla\widetilde{u}^{n+1}\|^2 \\
&+\tfrac{1}{\triangle t}\{\|\delta_t w^{n+1}\|^2 - \|\delta_t w^n\|^2 + \|\delta_t w^{n+1} - \delta_t w^n\|^2\} \\
&+\tfrac{1}{\triangle t}\{\|\nabla w^{n+1}\|^2 - \|\nabla w^n\|^2 + \triangle t^2\|\nabla\delta_t w^{n+1}\|^2\} + \triangle t\{\|\nabla p^{n+1}\|^2 - \|\nabla p^n\|^2\} = 0,
\end{aligned}$$

which implies the desired result. □

We recall that due to the artificial Dirichlet boundary condition for the pressure in (5.3.2c), a higher-order discretization for the velocity will not increase the accuracy. Hence, in order to obtain a higher-order scheme, one needs to resort to the rotational pressure-correction (cf. [79]).

5.3.2 Rotational Pressure-Correction Schemes

First-Order Scheme

We start by constructing a first-order scheme.

Step 1: Given (u^n, v^n, w^n, p^n), compute $\widetilde{u}^{n+1} \in \mathbf{H}^1_{0,\Gamma_f}(\Omega_f)$ and $w^{n+1} \in \mathbf{H}^1_{0,\Gamma_s}(\Omega_s)$ by solving

$$\begin{aligned}
&(\tfrac{\widetilde{u}^{n+1}-u^n}{\triangle t}, \varphi)_{\Omega_f} + (\mu\nabla\widetilde{u}^{n+1}, \nabla\varphi)_{\Omega_f} + \tfrac{1}{2}((u^n\cdot\mathbf{n})\widetilde{u}^{n+1}, \varphi)_{\Gamma_c} \\
&+((u^n\cdot\nabla)\widetilde{u}^{n+1}, \varphi)_{\Omega_f} - (p^n, \mathrm{div}\varphi)_{\Omega_f} + (\tfrac{\partial w^{n+1}}{\partial\mathbf{n}}, \varphi)_{\Gamma_c} = 0 \quad \forall\varphi \in \mathbf{H}^1_{0,\Gamma_f}(\Omega_f)
\end{aligned} \tag{5.3.8a}$$

$$\widetilde{u}^{n+1} = \frac{w^{n+1} - w^n}{\triangle t} \quad on \quad \Gamma_c \tag{5.3.8b}$$

$$(\frac{w^{n+1} - 2w^n + w^{n-1}}{\triangle t^2}, \psi)_{\Omega_s} + (\nabla w^{n+1}, \nabla\psi)_{\Omega_s} - (\frac{\partial w^{n+1}}{\partial\mathbf{n}}, \psi)_{\Gamma_c} = 0 \quad \forall\psi \in \mathbf{H}^1_{0,\Gamma_s}(\Omega_s) \tag{5.3.8c}$$

Step 2: Compute $u^{n+1} \in \mathbf{H}^1(\Omega_f)$ and $p^{n+1} \in H^1(\Omega_f)$ by solving

$$\begin{aligned}
&\tfrac{(u^{n+1}-\widetilde{u}^{n+1})}{\triangle t} + \nabla(p^{n+1} - p^n + \lambda\mu\mathrm{div}\tilde{u}^{n+1}) = 0, \quad in \quad \Omega_f \\
&\mathrm{div}\ u^{n+1} = 0, \quad in \quad \Omega_f \\
&u^{n+1}\cdot\mathbf{n}|_{\Gamma_f} = 0 \text{ and } p^{n+1}|_{\Gamma_c} = (p^n - \lambda\mu\mathrm{div}\tilde{u}^{n+1})|_{\Gamma_c}
\end{aligned} \tag{5.3.9a}$$

where $\lambda \in (0, \frac{2}{d})$ (with d being the space dimension) is a preselected parameter. We note that when $\lambda = 0$, the scheme reduces to the standard pressure-correction scheme.

We observe that main difference of the rotational scheme (5.3.8)-(5.3.9) with the standard scheme (5.3.1)-(5.3.3) is the additional term $\lambda\mu \mathrm{div}\tilde{u}^{n+1}$ in (5.3.9a). This term replace the artificial Dirichlet B.C. $p^{n+1}|_{\Gamma_c} = p^n|_{\Gamma_c}$ by an improved B.C. $p^{n+1}|_{\Gamma_c} = (p^n - \lambda\mu \mathrm{div}\tilde{u}^{n+1})|_{\Gamma_c}$. On the other hand, the numerical procedure for the two schemes are essentially identical.

Second Order Scheme

We also observe that it is not straightforward to construct a second-order version of (5.3.8)-(5.3.9) using the usual backward difference formula (BDF). Hence, we first introduce an additional variable $v = w_t$ and rewrite the FSI equations as

$$u_t - \mu\Delta u + (u \cdot \nabla)u + \nabla p = 0 \qquad in \quad \Omega_f \times (0,T) \tag{5.3.10a}$$

$$\mathrm{div} u = 0 \qquad in \quad \Omega_f \times (0,T) \tag{5.3.10b}$$

$$v_t - \Delta w = 0 \qquad in \quad \Omega_s \times (0,T) \tag{5.3.10c}$$

$$w_t - v = 0 \qquad in \quad \Omega_s \times (0,T) \tag{5.3.10d}$$

with the boundary condition

$$u = 0 \qquad on \quad \Gamma_f \times (0,T) \tag{5.3.11a}$$

$$w = 0 \qquad on \quad \Gamma_s \times (0,T) \tag{5.3.11b}$$

$$u = v \qquad on \quad \Gamma_c \times (0,T) \tag{5.3.11c}$$

$$\frac{\partial w}{\partial \mathbf{n}} = \mu \frac{\partial u}{\partial \mathbf{n}} - p\mathbf{n} - \frac{1}{2}(u \cdot \mathbf{n})u \qquad on \quad \Gamma_c \times (0,T) \tag{5.3.11d}$$

and the initial condition

$$u(\cdot, 0) = u_0 \qquad in \quad \Omega_f \tag{5.3.12a}$$

$$w(\cdot, 0) = w_0 \qquad in \quad \Omega_s \tag{5.3.12b}$$

$$v(\cdot, 0) = w_1 \qquad in \quad \Omega_s \tag{5.3.12c}$$

We can now construct a second-order rotational pressure-correction scheme as follows:

Step 1: Given (u^n, w^n,v^n, and p^n), compute $\widetilde{u}^{n+1} \in \mathbf{H}^1_{0,\Gamma_f}(\Omega_f)$ and $v^{n+1}, w^{n+1} \in \mathbf{H}^1_{0,\Gamma_s}(\Omega_s)$ by solving

$$\begin{aligned}&(\tfrac{3\widetilde{u}^{n+1}-4u^n+u^{n-1}}{2\triangle t},\varphi)_{\Omega_f}+(\mu\nabla\widetilde{u}^{n+1},\nabla\varphi)_{\Omega_f}+\tfrac{1}{2}(((2u^n-u^{n-1})\cdot\mathbf{n})\widetilde{u}^{n+1},\varphi)_{\Gamma_c}\\&+((2u^n-u^{n-1})\cdot\nabla\widetilde{u}^{n+1},\varphi)_{\Omega_f}-(p^n,\mathrm{div}\varphi)_{\Omega_f}+(\tfrac{\partial w^{n+1}}{\partial\mathbf{n}},\varphi)_{\Gamma_c}=0\quad\forall\varphi\in\mathbf{H}^1_{0,\Gamma_f}(\Omega_f)\end{aligned} \tag{5.3.13a}$$

$$\widetilde{u}^{n+1}=v^{n+1}\quad on\quad\Gamma_c \tag{5.3.13b}$$

$$\frac{3w^{n+1}-4w^n+w^{n-1}}{2\triangle t}-v^{n+1}=0\quad in\quad\Omega_s \tag{5.3.13c}$$

$$(\frac{3v^{n+1}-4v^n+v^{n-1}}{2\triangle t},\psi)_{\Omega_s}+(\nabla w^{n+1},\nabla\psi)_{\Omega_s}-(\frac{\partial w^{n+1}}{\partial\mathbf{n}},\psi)_{\Gamma_c}=0\quad\forall\psi\in\mathbf{H}^1_{0,\Gamma_s}(\Omega_f) \tag{5.3.13d}$$

Step 2: Compute (u^{n+1}, p^{n+1}) by solving

$$\begin{aligned}&\tfrac{3(u^{n+1}-\widetilde{u}^{n+1})}{2\triangle t}+\nabla(p^{n+1}-p^n+\lambda\mu\mathrm{div}\tilde{u}^{n+1})=0,\quad in\quad\Omega_f\\&\mathrm{div}\ u^{n+1}=0,\quad in\quad\Omega_f\\&u^{n+1}\cdot\mathbf{n}|_{\Gamma_f}=0\text{ and }p^{n+1}|_{\Gamma_c}=(p^n-\lambda\mu\mathrm{div}\tilde{u}^{n+1})|_{\Gamma_c}\end{aligned} \tag{5.3.14a}$$

where $\lambda \in (0, \frac{2}{d})$ is a preselected parameter.

Several remarks are in order:

- One observes that all the terms, except the pressure, are discretized with a second-order BDF or Adam-Bashforth formula. We recall that a first-order treatment of the pressure term, coupled with second-order treatment for other terms, may lead to second-order accuracy for the velocity [36].

- It is clear that, at each time step, the numerical procedure for solving (5.3.13)-(5.3.14) is essentially the same as solving the first-order scheme (5.3.1)-(5.3.3).

- The proof of unconditional stability for the rotational scheme is much more difficult. The original stability proof of the rotational pressure-correction scheme

in [80] was carried out only for Stokes problems, and an essential step of the proof was to take the "discrete time derivative" of the scheme. Unfortunately, this proof can not be extended to the nonlinear case.

- In [82], the author proved the unconditional stability for a Gauge-Uzawa scheme of the Navier-Stokes equations. A useful idea in [82] is to introduce a sequence $\{q^n\}$ defined by

$$q^n = \lambda\mu \mathrm{div}\tilde{u}^n + q^{n-1} \text{ with } q^{-1} = q^0 = 0. \tag{5.3.15}$$

We shall also use this sequence in our stability proof below.

Theorem 5.3.2 The scheme (5.3.13)-(5.3.14), with $p^{-1}|_{\Gamma_c} = p^0|_{\Gamma_c} = 0$, is unconditionally stable. More precisely, if we define the discrete energy as

$$\begin{aligned} E^{n+1} =& \|u^{n+1}\|^2 + \|2u^{n+1} - u^n\|^2 + \|v^{n+1}\|^2 + \|2v^{n+1} - v^n\|^2 + \|\nabla w^{n+1}\|^2 \\ &+ \|2\nabla w^{n+1} - \nabla w^n\|^2 + 2\triangle t\|q^{n+1}\|^2 + \frac{4\triangle t^2}{3}\|\nabla(p^{n+1} + q^{n+1})\|^2, \end{aligned} \tag{5.3.16}$$

then we have

$$E^{n+1} + \triangle t^4\|\delta_{tt}u^{n+1}\|^2 + \triangle t^4\|\delta_{tt}v^{n+1}\|^2 + \triangle t^4\|\delta_{tt}w^{n+1}\|^2 + (1-\lambda)4\triangle t\mu\|\nabla\tilde{u}^{n+1}\|^2 \le E^n.$$

Proof For any sequence $\{u^n, \tilde{u}^n\}$, we have

$$\begin{aligned} &(\frac{3\tilde{u}^{n+1} - 4u^n + u^{n-1}}{2\triangle t}, 4\triangle t\tilde{u}^{n+1})_{\Omega_f} = 2(3\tilde{u}^{n+1} - 4u^n + u^{n-1}, \tilde{u}^{n+1})_{\Omega_f} \\ =&6(\tilde{u}^{n+1} - u^{n+1}, \tilde{u}^{n+1})_{\Omega_f} + 2(3u^{n+1} - 4u^n + u^{n-1}, \tilde{u}^{n+1} - u^{n+1})_{\Omega_f} \\ &+ 2(3u^{n+1} - 4u^n + u^{n-1}, u^{n+1})_{\Omega_f} \end{aligned} \tag{5.3.17}$$

Let $I_1^n(u)$, $I_2^n(u)$ and $I_3^n(u)$ be the last three terms in the right-hand side. Using the algebraic identities

$$2(a^{k+1}, a^{k+1} - a^k) = |a^{k+1}|^2 - |a^k|^2 + |a^{k+1} - a^k|^2 \tag{5.3.18}$$

and

$$\begin{aligned} &2(a^{k+1}, 3a^{k+1} - 4a^k + a^{k-1}) \\ =&|a^{k+1}|^2 + |2a^{k+1} - a^k|^2 + |a^{k+1} - 2a^k + a^{k-1}|^2 - |a^k|^2 - |2a^k - a^{k-1}|^2, \end{aligned} \tag{5.3.19}$$

we find

$$I_1^n(u) = 3\|\tilde{u}^{n+1}\|^2 - 3\|u^{n+1}\|^2 + 3\|\tilde{u}^{n+1} - u^{n+1}\|^2,$$
$$I_3^n(u) = \|u^{n+1}\|^2 + \|2u^{n+1} - u^n\|^2 + \|u^{n+1} - 2u^n + u^{n-1}\|^2 - \|u^n\|^2 - \|2u^n - u^{n-1}\|^2. \tag{5.3.20}$$

Using the first equation in (5.3.14a), we have

$$I_2^n(u) = -\frac{4\triangle t}{3}(3u^{n+1} - 4u^n + u^{n-1}, \nabla(p^{n+1} - p^n + \lambda\mu \mathrm{div}\tilde{u}^{n+1}))_{\Omega_f} = 0.$$

Taking $\varphi = 4\triangle t\tilde{u}^{n+1}$ in (5.3.13a), and using (5.2.8) and the above relation, we obtain

$$I_1^n(u) + I_3^n(u) + 4\triangle t\mu\|\nabla\tilde{u}^{n+1}\|^2 - 4\triangle t(p^n, \mathrm{div}\ \tilde{u}^{n+1})_{\Omega_f} + 4\triangle t(\frac{\partial w^{n+1}}{\partial n}, \tilde{u}^{n+1})_{\Gamma_c} = 0. \tag{5.3.21}$$

Taking $\psi = 4\triangle t v^{n+1}$ in (5.3.13d), using (5.3.13b) and (5.3.13c), we find

$$I_3^n(v) + \tilde{I}_3^n(w) - 4\triangle t(\frac{\partial w^{n+1}}{\partial n}, \tilde{u}^{n+1})_{\Gamma_c} = 0, \tag{5.3.22}$$

where, by (5.3.20),

$$\begin{aligned}\tilde{I}_3^n(w) &= 2(\nabla(3w^{n+1} - 4w^n + w^{n-1}), \nabla w^{n+1})_{\Omega_f} \\ &= \|\nabla w^{n+1}\|^2 + \|2\nabla w^{n+1} - \nabla w^n\|^2 + \|\nabla w^{n+1} - 2\nabla w^n + \nabla w^{n-1}\|^2 \\ &\quad - \|\nabla w^n\|^2 - \|2\nabla w^n - \nabla w^{n-1}\|^2.\end{aligned} \tag{5.3.23}$$

Using (5.3.15), we can rewrite (5.3.14a) as

$$\frac{\sqrt{3}u^{n+1}}{\sqrt{\triangle t}} + \frac{2\sqrt{\triangle t}}{\sqrt{3}}\nabla(2p^{n+1} + q^{n+1}) = \frac{\sqrt{3}\widetilde{u}^{n+1}}{\sqrt{\triangle t}} + \frac{2\sqrt{\triangle t}}{\sqrt{3}}\nabla(p^n + q^n).$$

Taking the inner product with itself from both sides of the above equation, integrating parts and using (5.3.15) and the fact that $(p^k + q^k)|_{\Gamma_c} = \cdots = (p^0 + q^0)|_{\Gamma_c} = 0$, we obtain

$$\begin{aligned}&\frac{3}{\triangle t}\|u^{n+1}\|^2 + \frac{4\triangle t}{3}\|\nabla(p^{n+1} + q^{n+1})\|^2 - \frac{3\|\widetilde{u}^{n+1}\|^2}{\triangle t} - \frac{4\triangle t}{3}\|\nabla(p^n + q^n)\|^2 \\ &= -4(p^n + q^n, \mathrm{div}\ \widetilde{u}^{n+1})_{\Omega_f} = -4(p^n, \mathrm{div}\ \widetilde{u}^{n+1})_{\Omega_f} - \frac{4}{\lambda\mu}(q^n, q^{n+1} - q^n)_{\Omega_f} \\ &= -4(p^n, \mathrm{div}\ \widetilde{u}^{n+1})_{\Omega_f} + \frac{2}{\lambda\mu}\{\|q^n\|^2 - \|q^{n+1}\|^2 + \|q^{n+1} - q^n\|^2\}.\end{aligned} \tag{5.3.24}$$

Multiplying the above by Δt and adding it to (5.3.21), we obtain

$$\begin{aligned} 0 =& I_1^n(u) + I_3^n(u) + 4\triangle t\mu\|\nabla\tilde{u}^{n+1}\|^2 + I_3^n(v) + I_3^n(w) + 3\|u^{n+1}\|^2 \\ &+ \frac{4\triangle t^2}{3}\|\nabla(p^{n+1}+q^{n+1})\|^2 - 3\|\widetilde{u}^{n+1}\|^2 - \frac{4\triangle t^2}{3}\|\nabla(p^n+q^n)\|^2 \\ &- \frac{2}{\lambda\mu}\triangle t\{\|q^n\|^2 - \|q^{n+1}\|^2 + \|q^{n+1}-q^n\|^2\}. \end{aligned} \tag{5.3.25}$$

Thanks to (5.3.15), we have

$$\frac{2}{\lambda\mu}\|q^{n+1}-q^n\|^2 = 2\lambda\mu\|\text{div }\tilde{u}^{n+1}\|^2 \le 2\lambda\mu d\|\nabla\tilde{u}^{n+1}\|^2$$

where we have used the well-known inequality $\|\text{div }\tilde{u}^{n+1}\|^2 \le d\|\nabla\tilde{u}^{n+1}\|^2$ with $d = 2$ or 3 being the space dimension.

Finally, using the above inequality, (5.3.20) and (5.3.23) in (5.3.25), we find

$$\begin{aligned} & E^{n+1} - E^n \\ =& -\|u^{n+1}-2u^n+u^{n+1}\|^2 - \|v^{n+1}-2v^n+v^{n+1}\|^2 - \|\nabla(w^{n+1}-2w^n+w^{n+1})\|^2 \\ & - 4\triangle t\mu\|\nabla\tilde{u}^{n+1}\|^2 + 2d\lambda\triangle t\mu\|\text{div }\tilde{u}^{n+1}\|^2 \\ \le& -\|u^{n+1}-2u^n+u^{n+1}\|^2 - \|v^{n+1}-2v^n+v^{n+1}\|^2 - \|\nabla(w^{n+1}-2w^n+w^{n+1})\|^2 \\ & - (2-d\lambda)2\triangle t\mu\|\nabla\tilde{u}^{n+1}\|^2. \end{aligned}$$

which implies the desired result. □

Remark 5.3.3 With the stability results established in this section, it is also possible to derive similar error estimates for these schemes as in [79].

5.4 Galerkin type Spatial Discretization and Implementation

We briefly describe a general procedure to implement the time discretization schemes constructed in the last section. Let $\mathbf{X}_h \subset \mathbf{H}^1_{0,\Gamma_f}(\Omega_f)$, $M_h \subset H^1(\Omega_f)$, $M_h^0 = \{q \in M_h \,|\, q|_{\Gamma_c} = 0\}$ and $\mathbf{W}_h \subset \mathbf{H}^1_{0,\Gamma_s}(\Omega_s)$ be some finite dimensional approximation spaces, with $(\mathbf{X}_h, M_h)$ preferably satisfies the Babuska-Brezzi inf-sup condition. We also denote $\mathbf{Y}_h = \mathbf{X}_h + \nabla M_h^0$.

To fix the idea, we take the scheme (5.3.8)-(5.3.9) as an example. The other schemes can be treated by using exactly the same procedure.

5.4.1 A general setup

A Galerkin approximation of the scheme (5.3.8)-(5.3.9) is as follows:

Step 1. Let $\widetilde{w}_h^{n+1} = \delta_t w_h^{n+1}$. Then we look for $(u_h^{n+1}, \widetilde{w}_h^{n+1}) \in \mathbf{X}_h \times \mathbf{W}_h$ such that

$$\alpha(\tilde{u}_h^{n+1}, \varphi_h)_{\Omega_f} + (\nabla \tilde{u}_h^{n+1}, \nabla \varphi_h)_{\Omega_f} + ((u_h^n \cdot \nabla)\widetilde{u}_h^{n+1}, \varphi_h)_{\Omega_f}$$
$$+ \frac{1}{2}((u_h^n \cdot \mathbf{n})\widetilde{u}_h^{n+1}, \varphi_h)_{\Gamma_c} + \beta(\frac{\partial \tilde{w}_h^{n+1}}{\partial \mathbf{n}}, \varphi_h)_{\Gamma_c} =< f_h^n, \varphi_h >_{\Omega_f}, \quad \forall \varphi_h \in \mathbf{X}_h, \tag{5.4.1a}$$

$$\tilde{u}_h^{n+1} = \tilde{w}_h^{n+1}, \quad \text{at } \Gamma_c, \tag{5.4.1b}$$

$$\alpha(\tilde{w}^{n+1}, \psi_h)_{\Omega s} + \beta(\nabla \tilde{w}^{n+1}, \nabla \psi_h)_{\Omega_s} - \beta(\frac{\partial \tilde{w}^{n+1}}{\partial \mathbf{n}}, \psi_h)_{\Gamma_c} =< g_h, \psi_h >_{\Omega_s}, \quad \forall \psi_h \in \mathbf{W}_h, \tag{5.4.1c}$$

where $\alpha = \frac{1}{\triangle t}, \beta = \triangle t$, and

$$< f_h^n, \varphi_h >_{\Omega_f} := \alpha(u_h^n, \varphi_h)_{\Omega_f} + (p_h^n, \mathrm{div}\varphi_h)_{\Omega_f} - (\frac{\partial w_h^n}{\partial \mathbf{n}}, \varphi_h)_{\Gamma_c}, \tag{5.4.2}$$

and

$$< g_h^n, \psi_h >_{\Omega_s} := \alpha(\widetilde{w}_h^n, \psi_h^n)_{\Omega_s} - (\nabla w_h^n, \nabla \psi_h)_{\Omega_s} + (\frac{\partial w_h^n}{\partial \mathbf{n}}, \psi_h)_{\Gamma_c}. \tag{5.4.3}$$

Define

$$\hat{u}_h^{n+1}(x,y) = \begin{cases} \tilde{u}_h^{n+1}(x,y), & \text{if } (x,y) \in \Omega_f, \\ \tilde{w}_h^{n+1}(x,y), & \text{if } (x,y) \in \Omega_s; \end{cases}$$

$$\hat{\beta}(x,y) := \begin{cases} 1, & \text{if } (x,y) \in \Omega_f, \\ \beta, & \text{if } (x,y) \in \Omega_s. \end{cases}$$

and

$$b(u, v, \varphi) := ((u \cdot \nabla)\widetilde{v}, \varphi)_{\Omega_f} + (\frac{1}{2}(u \cdot \mathbf{n})\widetilde{v}, \varphi)_{\Gamma_c}.$$

Then, we can rewrite (5.4.1) as: Find $\hat{u}_h^{n+1} \in \mathbf{X}_h \times \mathbf{W}_h \cap \mathbf{H}^1(\Omega)$ such that

$$\begin{aligned} \alpha(\hat{u}_h^{n+1}, \phi_h) + (\hat{\beta}\nabla \hat{u}_h^{n+1}, \nabla \phi_h) + b(u_h^n, \hat{u}_h^{n+1}, \phi_h) \\ =< f_h^n, \phi_h >_{\Omega_f} + < g_h^n, \phi_h >_{\Omega_s}, \quad \forall \phi_h \in \mathbf{X}_h \times \mathbf{W}_h \cap \mathbf{H}^1(\Omega). \end{aligned} \tag{5.4.4}$$

Thus, the equation (5.4.4) can be viewed as a domain-decomposition (with two-domains) approximation to a linear elliptic problem with discontinuous coefficient

$\hat{\beta}$. Note that from (5.4.1b), $\hat{u}_h^{n+1}(x,y)$ is continuous at Γ_c. Hence, one can efficiently solve the coupled linear system using a standard domain decomposition approach, in particular, in the two-dimension case, one can form the Schur-complement to solve the unknown at the interface first, and then solve for the velocity in the fluid region and displacement in the solid region separately (cf., for instance, [86,87] and a simple example in the next subsection).

Step 2. Find $\phi_h^{n+1} \in M_h^0$ such that

$$(\nabla\phi_h^{n+1}, \nabla q_h)_{\Omega_f} = \frac{1}{\triangle t}(\widetilde{u}_h^{n+1}, \nabla q_h)_{\Omega_f}, \quad \forall q_h \in M_h^0; \tag{5.4.5}$$

and compute $u_h^{n+1} \in \mathbf{Y}_h$ and $p_h^{n+1} \in M_h$ by

$$\begin{aligned} u_h^{n+1} &= \widetilde{u}_h^{n+1} - \triangle t\nabla\phi_h^{n+1}, \\ p_h^{n+1} &= p_h^n + \phi_h^{n+1} - \lambda\mu Q_h \mathrm{div}\widetilde{u}_h^{n+1}, \end{aligned} \tag{5.4.6}$$

where Q_h is a L^2-projection operator onto M_h.

We note that (5.4.5) is just a discrete Poisson equation in Ω_f with homogeneous Dirichlet boundary condition on Γ_c, and (5.4.6) involves only a projection, so they can be efficiently solved.

5.4.2 An example with a Fourier-Legendre approximation

As an example, we consider a two-dimensional periodic channel with $\Omega_f = (0, 2\pi)\times(0,1)$, $\Omega_s = (0,2\pi)\times(-1,0)$, so $\Omega = (0,2\pi)\times(-1,1)$, $\Gamma_f = \{(x,y)|\, x \in (0,2\pi), y = 1\}$, $\Gamma_c = \{(x,y)|\, x \in (0,2\pi), y = 0\}$ and $\Gamma_s = \{(x,y)|\, x \in (0,2\pi), y = -1\}$. We denote $\mathbf{I}^+, \mathbf{I}^-, \mathbf{I}$ by $\mathbf{I}^+ = [0,1]$, $\mathbf{I}^- = [-1,0]$ and $\mathbf{I} = [-1,1]$. We assume that all functions are periodic in the x-direction.

Let $h = (M, N)$ where M is the number of equally spaced points in the x-direction, and $N+1$ is the number of Legendre-Gauss-Lobatto points in the y direction of Ω_f and Ω_s. For simplicity, we have assumed to use the same number of points in the y

direction of Ω_f and Ω_s, while in practice, different number of points can be used. Let P_N be the set of all polynomials of degree less than or equal to N. We set

$$
\begin{aligned}
&X_h = \{v_h = \sum_{k=-M/2}^{M/2} v_k(y)e^{ikx} \text{ with } v_k(\cdot) \in P_N,\ v_k(1) = 0\},\ \mathbf{X}_h = X_h \times X_h,\\
&W_h = \{w_h = \sum_{k=-M/2}^{M/2} w_k(y)e^{ikx} \text{ with } w_k(\cdot) \in P_N,\ w_k(-1) = 0\},\ \mathbf{W}_h = W_h \times W_h,\\
&M_h^0 = \{q_h = \sum_{k=-M/2}^{M/2} q_k(y)e^{ikx} \text{ with } q_k(\cdot) \in P_{N-1}, q_k(0) = 0\},\ \mathbf{Y}_h = \mathbf{X}_h + \nabla M_h^0,\\
&X_N^0 = \{v \in H^1(I) : v|_{\mathbf{I}^+}, v|_{\mathbf{I}^-} \in P_N,\ v(-1) = v(1) = 0\},\ \mathbf{X}_N^0 = X_N^0 \times X_N^0.
\end{aligned}
\tag{5.4.7}
$$

For the sake of efficiency and to take full advantage of periodicity, we shall treat the nonlinear convective term in (5.4.1) expplilicitly. To this end, we modify (5.4.2) to

$$
< f_h^n, \varphi_h >_{\Omega_f} := \alpha(u_h^n, \varphi_h)_{\Omega_f} + (p_h^n, \mathrm{div}\varphi_h)_{\Omega_f} - (\frac{\partial w_h^n}{\partial \mathbf{n}}, \varphi_h)_{\Gamma_c} - b(u_h^n, u_h^n, \varphi_h). \tag{5.4.8}
$$

With this modification and expand all the functions in discrete Fourier series, e.g.,

$$
\left(\hat{u}_h^{n+1}, f_h^n, g_h^n\right) = \sum_{m=-M/2}^{M/2} \left(u_m^{n+1}(y), f_m^n(y), g_m^n(y)\right) e^{imx}, \tag{5.4.9}
$$

we find that (5.4.4) reduces to: Find $u_m^{n+1} \in \mathbf{X}_N^0$ such that

$$
(\alpha_m u_m^{n+1}, \phi)_{\mathbf{I}} + (\hat{\beta}\frac{du_m^{n+1}}{dy}, \frac{d\phi}{dy})_{\mathbf{I}} = (f_m^n, \phi)_{\mathbf{I}^+} + (g_m^n, \phi)_{\mathbf{I}^-}, \quad \forall \phi \in \mathbf{X}_N^0, \tag{5.4.10}
$$

where

$$
\alpha_m = \begin{cases} \alpha + m^2, & \text{if } y \in \mathbf{I}^+, \\ \alpha + \beta m^2, & \text{if } y \in \mathbf{I}^-. \end{cases}
$$

Next we construct a set of basis functions for X_N^0.

We define, for $i = 0, 1, ..., N-2$,

$$
\hat{\varphi}_i(y) = \begin{cases} L_k(2y-1) - L_{k+2}(2y-1), & \text{if } y \in \mathbf{I}^+, \\ 0, & \text{if } y \in \mathbf{I}^-; \end{cases}
$$

$$\hat{\varphi}_{N-1+i}(y) = \begin{cases} 0, & \text{if } y \in \mathbf{I}^+, \\ L_k(1+2y) - L_{k+2}(1+2y), & \text{if } y \in \mathbf{I}^-; \end{cases}$$

and the basis function at the interface is

$$\hat{\varphi}_{2N-2} = \begin{cases} 1-y, & \text{if } y \in \mathbf{I}^+, \\ 1+y, & \text{if } y \in \mathbf{I}^-. \end{cases}$$

Then

$$X_N^0 = \text{span}\left\{\hat{\varphi}_0, \hat{\varphi}_1, \cdots, \hat{\varphi}_{2N-2}\right\}. \tag{5.4.11}$$

Then, writing

$$u_m^{n+1}(y) = \sum_{k=0}^{2N-2} \hat{u}_{m,k}^{n+1}\hat{\varphi}_k(y),\ \hat{f}_{m,k}^n = (f_m^n, \hat{\varphi}_k)_{\mathbf{I}^+} + (g_m^n, \hat{\varphi}_k)_{\mathbf{I}^-},$$

and take $\phi = \hat{\varphi}_k$ in (5.4.10), we can derive the following linear system:

$$\left(\alpha \begin{bmatrix} M_{11} & 0 & m_{13} \\ 0 & M_{22} & m_{23} \\ m_{31}^T & m_{32}^T & m_{33} \end{bmatrix} + \begin{bmatrix} S_{11} & 0 & s_{13} \\ 0 & S_{22} & s_{23} \\ s_{31}^T & s_{32}^T & s_{33} \end{bmatrix}\right) \begin{bmatrix} \bar{u}_1 \\ \bar{u}_2 \\ \bar{u}_3 \end{bmatrix} = \begin{bmatrix} \bar{f}_1 \\ \bar{f}_2 \\ \bar{f}_3 \end{bmatrix}, \tag{5.4.12}$$

where $\bar{u}_1 = (\hat{u}_{m,0}^{n+1}, \hat{u}_{m,1}^{n+1}, \cdots, \hat{u}_{m,N-2}^{n+1})^T$, $\bar{u}_2 = (\hat{u}_{m,N-1}^{n+1}, \hat{u}_{m,N}^{n+1}, \cdots, \hat{u}_{m,2N-3}^{n+1})^T$ and $\bar{u}_3 = u_{m,2N-2}^{n+1}$, similarly for $\bar{f}_1$, $\bar{f}_2$ and $\bar{f}_3$; M_{ij} and S_{ij} are block mass and stiffness matrices. We recall that M_{ii} $(i = 1, 2)$ are penta-diagonal and S_{ii} $(i = 1, 2)$ are diagonal (cf. [88, 89]). So the linear system can be easily solved by the Schur-complement approach, More precisely, solve first $\bar{u}_3$ using a block Gaussian elimination, and then solve $\bar{u}_1$ and $\bar{u}_2$ separately.

5.5 Numerical Results

To examine the correctness and accuracy of the proposed numerical schemes, we consider the following non-homogeneous problem

$$u_t - \Delta u + (u \cdot \nabla)u + \nabla p = f \qquad in \quad \Omega_f \times (0, T) \tag{5.5.1a}$$

$$\text{div} u = 0 \qquad in \quad \Omega_f \times (0, T) \tag{5.5.1b}$$

$$w_{tt} - \Delta w = g \qquad in \quad \Omega_s \times (0, T) \tag{5.5.1c}$$

with the boundary condition:

$$u = 0 \qquad on \quad \Gamma_f \times (0,T) \tag{5.5.2a}$$

$$w = 0 \qquad on \quad \Gamma_s \times (0,T) \tag{5.5.2b}$$

$$u = w_t \qquad on \quad \Gamma_c \times (0,T) \tag{5.5.2c}$$

$$\frac{\partial w}{\partial n} = \frac{\partial u}{\partial n} - p\mathbf{n} - \frac{1}{2}(u \cdot \mathbf{n})u + h \qquad on \quad \Gamma_c \times (0,T) \tag{5.5.2d}$$

where $\Omega_f = (0, 2\pi) \times (0, 1)$, $\Omega_s = (0, 2\pi) \times (-1, 0)$ with periodic boundary conditions in the x-direction.

We set the exact solution to be

$$\begin{aligned} u &= (-\sin(\pi t)\cos(x)\sin(y-1), \sin(\pi t)\sin(x)(\cos(y-1)-1)), \\ p &= \sin(\pi t)\cos(x)\cos(y), \\ w &= (-\cos(\pi t)\cos(x)\sin(y-1), -\cos(\pi t)\sin(x)(\cos(y+1)-1)). \end{aligned} \tag{5.5.3}$$

The functions f, g, h can then be computed accordingly.

We employ the Fourier-Legendre method presented in the last section, and choose (M, N) large enough so that the errors are dominated by that from the time discretization. In the following examples, we choose $\lambda = 0.5$, which is a preselected parameter introduced in (5.3.9a) and (5.3.14a).

In Figure 5.2, we plot the L^2-errors for the pressure and for the velocity and displacement with the second-order standard and rotational pressure-correction schemes. We observe that the rotational scheme perform much better than the standard scheme.

In Figure 5.3, we plot the convergence rate of the second-order rotational scheme. We consider ending time $T = 2$ and vary the step size from $\triangle t = 0.1$ to $\triangle t = 0.0001$. We observe that the L_2 errors for the fluid velocity, the structure displacement and the pressure all converge at a rate close to 3/2. We note that similar convergence rate was observed for Stokes equations with open boundary (cf. [79]).

Next, we examine the energy stability of our schemes by solving the homogeneous (with f, g and h being zero) FSI problem with the same initial conditions as in the last example. We take the second-order rotational scheme as an example, and plot

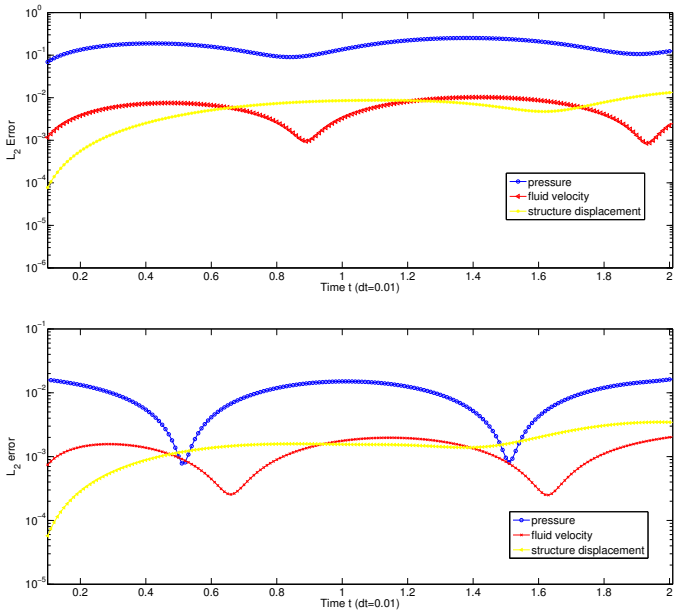


Figure 5.2. dt=0.01, T=2; 2nd order scheme; Top: standard; Bottom: rotational.

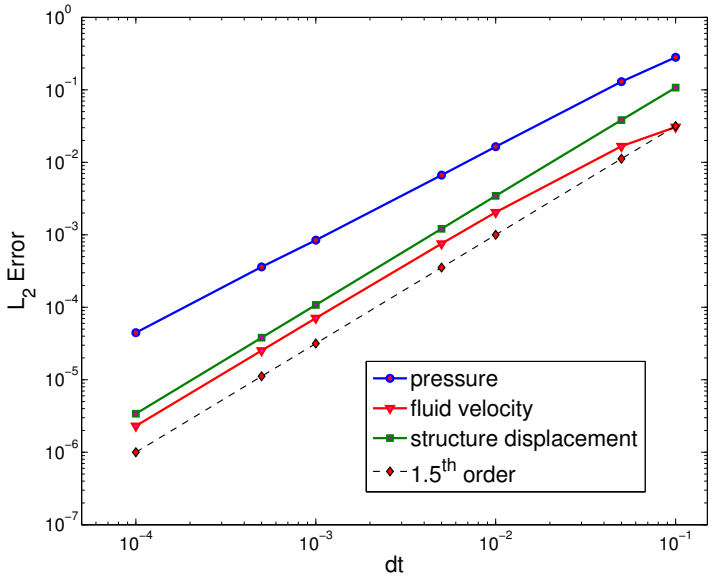


Figure 5.3. L_2 Error for second-order rotational scheme.

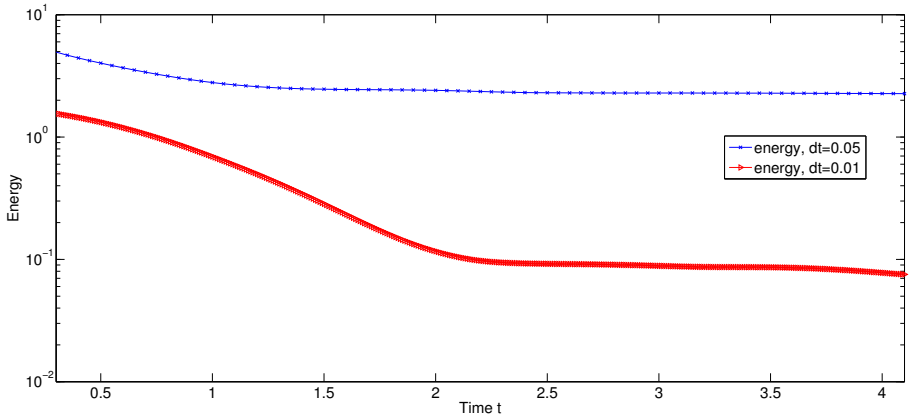


Figure 5.4. Energy decay for time step 0.01 and 0.05

in Figure 5.4 the discrete energy for the cases with $\triangle t = 0.01$ and $\triangle t = 0.05$. We observe that the discrete energy indeed decays monotonically.

5.6 Conclusion

We constructed in this paper standard and rotational pressure correction schemes for the FSI problem with a fixed interface, and proved rigorously that they are unconditionally energy stable. These schemes are computationally very efficient: at each time step, they lead to (i) a coupled linear elliptic system for the velocity and displacement, with the coupling condition at the interface between the fluid and solid regions, which can be efficiently solved by using a standard domain decomposition (with two domains) approach; and (ii) a discrete Poisson equation in the fluid region.

We validated these schemes by using a Fourier-Legendre spatial discretization for the FSI problem in a periodic channel.

Although we only considered the FSI problem with fixed interface, we believe that some of the essential approaches in constructing our numerical schemes can be extended to the FSI problems with moving interface which we plan to address in a future endeavor.

CHAPTER 6. EXTENSIONS AND FUTURE WORK

6.1 High-order Method for Scattering Problems from Open Cavity

The phenomenon of electromagnetic scattering by cavity-backed apertures has received much attention by both the engineering and mathematical communities for its important applications. In this chapter, we mainly focus on the numerical approximation for the Helmholze equation obtaining from the scattering problems from an single open cavity. Some preliminary numerical analysis has been made to the governing mathematical problem, but further implementation and numerical verification will be our future work.

6.1.1 A Model Problem

As shown in Figure 6.1, an open cavity Ω, enclosed by the aperture Γ and the wall S, is placed on a perfectly conducting ground plane Γ^{c}. Above the flat surface $\{y = 0\} = \Gamma \cup \Gamma^{\mathrm{c}}$, the medium is assumed to be homogeneous with a positive dielectric permittivity ε_0 and magnetic permeability μ_0. Inside the cavity Ω, it is assumed to be filled with some layered medium, which can be described by a y-dependent relative dielectric permittivity $\varepsilon(y)$.

We consider the two-dimensional Helmholtz equation

$$\Delta u + \kappa^2 u = 0, \quad \text{in } \Omega \cup \mathbb{R}^2_+, \tag{6.1.1}$$

together with the homogeneous Dirichlet boundary condition

$$u = 0, \quad \text{on } \Gamma^{\mathrm{c}} \cup S. \tag{6.1.2}$$

Here $\kappa^2 = \omega^2 \varepsilon \mu_0$, where ω is the angular frequency and κ is known as the wavenumber.

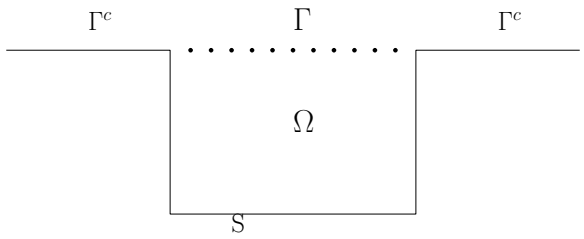


Figure 6.1. The problem geometry for a single cavity scattering problem. An open cavity Ω, enclosed by the aperture Γ and the wall S, is placed on a perfectly conducting ground plane Γ^c.

Let an incoming plane wave $u^{\mathrm{i}} = e^{\mathrm{i}\kappa_0(x\sin\theta - y\cos\theta)}$ be incident on the cavity from above, where $\theta \in (-\pi/2, \pi/2)$ is the angle of incidence with respect to the positive y-axis, and $\kappa_0 = \omega\sqrt{\varepsilon_0\mu_0}$ is the wavenumber of the free space.

Denote the reference field u^{ref} as the solution of the homogeneous Helmholtz equation in the upper half space:

$$\Delta u^{\mathrm{ref}} + \kappa_0^2 u^{\mathrm{ref}} = 0, \quad \text{in } \mathbb{R}^2_+, \tag{6.1.3}$$

together with the boundary condition

$$u^{\mathrm{ref}} = 0 \quad \text{on } \Gamma^{\mathrm{c}} \cup \Gamma. \tag{6.1.4}$$

To fix the idea, we set $\Gamma = (-1, 1)$.

It can be shown from (6.1.3) and (6.1.4) that the reference field consists of the incident field u^{i} and the reflected field u^{r}:

$$u^{\mathrm{ref}} = u^{\mathrm{i}} + u^{\mathrm{r}},$$

where $u^{\mathrm{r}} = -e^{\mathrm{i}\kappa_0(x\sin\theta + y\cos\theta)}$.

The total field u is composed of the reference field u^{ref} and the scattered field u^{s}:

$$u = u^{\mathrm{ref}} + u^{\mathrm{s}}.$$

It can be verified from (6.1.1) and (6.1.3) that the scattered field satisfies

$$\Delta u^{\mathrm{s}} + \kappa_0 u^{\mathrm{s}} = 0, \quad \text{in } \mathbb{R}^2_+. \tag{6.1.5}$$

In addition, the scattered field is required to satisfy the radiation condition

$$\lim_{\rho\to 0} \sqrt{\rho}\left(\frac{\partial u^{\mathrm{s}}}{\partial \rho} - \mathrm{i}\kappa_0 u^{\mathrm{s}}\right) = 0, \quad \rho = |(x, y)|. \tag{6.1.6}$$

6.1.2 Transparent Boundary Condition

By taking the Fourier transform of (6.1.5) with respect to x, we have an ordinary differential equation with respect to y:

$$\frac{\partial \hat{u}^{\mathrm{s}}(\xi, y)}{\partial y^2} + (\kappa_0^2 - \xi^2)\hat{u}^{\mathrm{s}}(\xi, y) = 0, \quad y > 0. \tag{6.1.7}$$

Since the solution of (6.1.7) satisfies the radiation condition (6.1.6), we deduce that the solution of (6.1.7) has the analytical form

$$\hat{u}^{\mathrm{s}}(\xi, y) = \hat{u}^{\mathrm{s}}(\xi, 0) e^{\mathrm{i}\beta(\xi) y}, \tag{6.1.8}$$

where

$$\beta(\xi) = \begin{cases} (\kappa_0^2 - \xi^2)^{1/2} & \text{for} \quad |\xi| < \kappa_0, \\ \mathrm{i}(\xi^2 - \kappa_0^2)^{1/2} & \text{for} \quad |\xi| > \kappa_0. \end{cases}$$

Taking the inverse Fourier transform of (6.1.8), we find that

$$u^{\mathrm{s}}(x, y) = \int_{\mathbb{R}} \hat{u}^{\mathrm{s}}(\xi, 0) e^{\mathrm{i}\beta(\xi) y} e^{-\mathrm{i}\xi x} d\xi \quad \text{in } \mathbb{R}^2_+.$$

Taking the normal derivative on $\Gamma^{\mathrm{c}} \cup \Gamma$, which is the partial derivative with respect to y on $\Gamma^{\mathrm{c}} \cup \Gamma$, and evaluating at $y = 0$ yield

$$\partial_y u^{\mathrm{s}}(x, y)\big|_{y=0} = \int_{\mathbb{R}} \mathrm{i}\beta(\xi) \hat{u}^{\mathrm{s}}(\xi, 0) e^{-\mathrm{i}\xi x} d\xi. \tag{6.1.9}$$

For any given u on $\Gamma^{\mathrm{c}} \cup \Gamma$, define the boundary operator T:

$$Tu = \int_{\mathbb{R}} \mathrm{i}\beta(\xi) \hat{u}(\xi, 0) e^{-\mathrm{i}\xi x} d\xi, \tag{6.1.10}$$

which leads to a transparent boundary condition for the scattered field on $\Gamma^{\mathrm{c}} \cup \Gamma$:

$$\partial_{\mathbf{n}}(u - u^{\mathrm{ref}}) = T(u - u^{\mathrm{ref}}).$$

Equivalently it can be written as a transparent boundary condition for the total field

$$\partial_{\mathbf{n}} u = Tu + g \quad \text{on } \Gamma^{\mathrm{c}} \cup \Gamma, \tag{6.1.11}$$

where

$$g(x) = \partial_{\mathbf{n}} u^{\mathrm{ref}} - Tu^{\mathrm{ref}} = -2\mathrm{i}\kappa_0 \cos\theta e^{\mathrm{i}\kappa_0 x \sin\theta}.$$

Then, we get the model problem for the total field u:

$$\Delta u + \kappa^2 u = 0, \quad \text{in } \Omega, \tag{6.1.12}$$

$$\partial_{\mathbf{n}} u = Tu + g, \quad \text{on } \Gamma, \tag{6.1.13}$$

$$u = 0, \quad \text{on } S. \tag{6.1.14}$$

6.1.3 Numerical Approximation by Legendre-Spectral Method

The system (6.1.12) - (6.1.14) looks very similar to scattering problem from periodic structure after taking the DtN technique, except that the periodic boundary condition is replaced by the Dirichlet boundary condition. Therefore, the TFE approach may not be able to directly applied here, but we can still follow the SEM approach used in Chapter 3.

A Directly Approach

To simplify the illustration on how to apply the SEM approach on the open cavity problem, we only use one element here as an example, which is also equivalent to the Spectral Method.

Assume the numerical solution has the form

$$u(x,y) = \sum_{m,n=0}^{N} u_{m,n}\psi_m(x)\phi_n(y) \tag{6.1.15}$$

where $\psi_m(x) = L_m(x) - L_{m+2}(x)$, $\phi_n(y) = L_n(y) + L_{n+1}(y)$ and$L_m(x), L_n(y)$ are Legendre polynomial basis.

Inserting into system (6.1.12) - (6.1.14), we have

$$\sum_{m,n=0}^{N} u_{m,n}\{\partial_{xx}\psi_m(x)\phi_n(y) + \psi_m(x)\partial_{yy}\phi_n(y) + \kappa^2\psi_m(x)\phi_n(y)\} = 0, \quad \text{in } \Omega, \tag{6.1.16}$$

$$\sum_{m,n=0}^{N} u_{m,n}\{\psi_m(x)\partial_{\mathbf{n}}\phi_n(1) - T[\psi_m(x)]\phi_n(1)\} = g, \quad \text{on } \Gamma. \tag{6.1.17}$$

which results in the matrix system

$$S_x U M_y^t + M_x U S_y^t + T_x U M_y^t(1) = G \tag{6.1.18}$$

where $M_x = (\psi_i(x), \psi_j(x))$, $S_x = -(\partial_x\psi_i(x), \partial_x\psi_j(x))$, $T_x = (T\psi_i(x), \psi_j(x))$, $S_y = -(\partial_y\phi_i(y), \partial_y\phi_j(y))$, $M_y = (\phi_i(y), \phi_j(y))$, $M_y = (\phi_i(1)\phi_j(1))_{ij}$ and $G = (g, \psi_m(x)\phi_n(y))$.

We can also rewrite the above matrix equation in the following form using the tensor product notation:

$$(S_x \bigotimes M_y^t + M_x \bigotimes S_y^t + T_x \bigotimes M_y^t(1))\bar{u} = \bar{g} \tag{6.1.19}$$

By the definition, S_x is diagonal matrix, M_x five-diagonal nonzero symmetric matrix, S_y is a diagonal matrix, M_y tri-diagonal nonzero matrix, but T_x, $M_y^t(1)$ are almost full dense matrices. To solve the matrix system fast and efficiently, if without T_x term, we can apply the Matrix diagonalization method ([73], [90]), however, with T_x terms, it makes the problem much more difficult as we have to solve a tensor product based big system. Therefore, the next approach will be focused on how to approximate the matrix T_x more accurately and efficiently.

Remark 6.1.1 :

$$\begin{aligned}
& (T_x)_{mn} && (6.1.20)\\
= \; & (T\psi_m(x), \psi_n(x)) && (6.1.21)\\
= \; & \int_R i\sqrt{\kappa^2 - x^2}\hat{\psi}_m(x)\hat{\psi}_n(-x)dx && (6.1.22)\\
= \; & \int_{-1}^{1}\int_{-1}^{1} \psi_m(x)\psi_n(y) \int_R i\sqrt{\kappa^2 - \xi^2}e^{i\xi(x-y)}d\xi dxdy && (6.1.23)
\end{aligned}$$

in either form, which requires a whole real line integration of the Fourier transform of the Legendre functions or $\sqrt{\kappa^2 - \xi^2}$. Therefore, a fast and accurate numerical integration for our problem is difficult and necessary.

Fourier Transform Approach

Use the similar idea in Chapter 4 for the unbounded rough surface scattering problem.

The basis function $\psi_m(\xi), m \geq 0$ satisfy the zero boundary condition at $x = \pm 1$, so we can do the zero extension and take fourier transform to (6.1.16):

$$\sum_{m,n=0}^{N} u_{m,n}\{-\xi^2\hat{\psi}_m(\xi)\phi_n(y) + \hat{\psi}_m(\xi)\partial_{yy}\phi_n(y) + \kappa^2\hat{\psi}_m(\xi)\phi_n(y)\} = 0, \quad \text{in } \Omega, \tag{6.1.24}$$

$$\sum_{m,n=0}^{N} u_{m,n}\{\hat{\psi}_m(\xi)\partial_{\mathbf{n}}\phi_n(1) - i\beta(\xi)\hat{\psi}_m(\xi)]\phi_n(1)\} = \hat{g}, \quad \text{on } \Gamma. \tag{6.1.25}$$

For zero ξ, as $\hat{\psi}_m(0) = \hat{L}_m(0) - \hat{L}_{m+2}(0) = \hat{L}_m(0) = \sqrt{\frac{2}{\pi}}\delta_{m,0}$, the system can be simplified and solved for $m = 0$ in a 1D problem w.r.t y:

$$\sum_{n=0}^{N} u_{0,n}\{\partial_{yy}\phi_n(y) + \kappa^2\phi_n(y)\} = 0, \quad y \in (-1, 1), \tag{6.1.26}$$

$$\sum_{n=0}^{N} u_{0,n}\{\partial_{\mathbf{n}}\phi_n(1) - i\kappa]\phi_n(1)\} = \sqrt{\frac{\pi}{2}}\hat{g}, \quad y = 1. \tag{6.1.27}$$

Therefore, we can obtain $\{u_{0,n}\}_{n=0}^{N}$ by solving above equations.

After obtaining $\{u_{0,n}\}_{n=0}N$, we can move the terms including $\{u_{0,n}\}_{n=0}N$ to the right hand side of the equations. Therefor, for nonzero ξ, we can multiply $\frac{1}{\xi^2}$ to (6.1.24) and (6.1.24), and then take inner product with $\hat{\psi}_m(\xi), m \geq 1$ and obtain

$$\sum_{m=1,n=0}^{N} u_{m,n}\{\frac{1}{\xi^2}\hat{\psi}_m(\xi)\partial_{yy}\phi_n(y) + (\frac{\kappa^2}{\xi^2} - 1)\hat{\psi}_m(\xi)\phi_n(y)\} = \frac{1}{\xi^2}f(\xi, y) \tag{6.1.28}$$

$$\sum_{m,n=0}^{N} u_{m,n}\{\frac{1}{\xi^2}\hat{\psi}_m(\xi)\partial_{\mathbf{n}}\phi_n(1) - \frac{i\beta(\xi)}{\xi^2}\hat{\psi}_m(\xi)\phi_n(1)\} = \frac{1}{\xi^2}(\hat{g} - h(\xi, y)) \tag{6.1.29}$$

$$f(\xi, y) = -\sum_{n=0}^{N} u_{0,n}\hat{\psi}_0(\xi)\partial_{yy}\phi_n(y) + (\kappa^2 - \xi^2)\hat{\psi}_0(\xi)\phi_n(y)\}, \quad \text{in } R \times [-1, 1],$$

$$h(\xi, y) = \sum_{n=0}^{N} u_{0,n}\{\hat{\psi}_0(\xi)\partial_{\mathbf{n}}\phi_n(1) - i\beta(\xi)\hat{\psi}_0(\xi)]\phi_n(1)\}, \quad \text{on } R.$$

Noticed that

$$(\frac{1}{\xi^2}\hat{\psi}_m(\xi), \hat{\psi}_n(\xi)) = (\frac{1}{\xi}\hat{\psi}_m(\xi), \frac{1}{\xi}\hat{\psi}_n(\xi)) \tag{6.1.30}$$

$$(\hat{\psi}_m(\xi), \hat{\psi}_n(\xi)) = (\psi_m(x), \psi_n(x)) = M_x \tag{6.1.31}$$

$$(\frac{i\beta(\xi)}{\xi^2}\hat{\psi}_m(\xi),\hat{\psi}_n(\xi)) = (i\sqrt{\frac{1}{\xi^2}(\frac{\kappa^2}{\xi^2}-1)}\hat{\psi}_m(\xi),\hat{\psi}_n(\xi)) \tag{6.1.32}$$

which results in a similar matrix system

$$\bar{M}_x U S_y^t + (\kappa^2\bar{M}_x - \bar{S}_x)UM_y^t + \bar{T}_x U M_y^t(1) = \bar{G} \tag{6.1.33}$$

where $\bar{M}_x = (\frac{1}{\xi^2}\hat{\psi}_i(\xi),\hat{\psi}_j(\xi))$, $\bar{S}_x = (\hat{\psi}_i(\xi),\hat{\psi}_j(\xi))$, $\bar{T}_x = (\frac{i\beta(\xi)}{\xi^2}\hat{\psi}_m(\xi),\hat{\psi}_n(\xi))$, $S_y = -(\partial_y\phi_i(y),\partial_y\phi_j(y))$, $M_y = (\phi_i(y),\phi_j(y))$, $M_y = (\phi_i(1)\phi_j(1))_{ij}$, and $\bar{G} = (\frac{1}{\xi^2}\hat{g},\hat{\psi}_m(\xi)\phi_n(y))$.

By further calculation using the property in next section, we have $\bar{S}_x$ is diagonal matrix, $\bar{M}_x$ five-diagonal nonzero symmetric matrix, S_y is a diagonal matrix, M_y tri-diagonal nonzero matrix, but $\bar{T}_x$, $M_y^t(1)$ are still full dense matrices.

Remark 6.1.2 :

$$\begin{aligned}
& (\bar{T}_x)_{mn} && (6.1.34)\\
= \ & (\frac{i\beta(\xi)}{\xi^2}\hat{\psi}_m(\xi),\hat{\psi}_n(\xi)) && (6.1.35)\\
= \ & \int_R i\sqrt{\frac{1}{\xi^2}(\frac{\kappa^2}{\xi^2}-1)}\hat{\psi}_m(\xi)\hat{\psi}_n(-\xi)d\xi && (6.1.36)\\
= \ & \int_{-1}^1\int_{-1}^1 \psi_m(x)\psi_n(y)\int_R i\sqrt{\frac{1}{\xi^2}(\frac{\kappa^2}{\xi^2}-1)}e^{i\xi(x-y)}d\xi dxdy && (6.1.37)
\end{aligned}$$

in either form, which requires an integration of the Fourier transform of the Legendre functions or $\sqrt{\frac{1}{\xi^2}(\frac{\kappa^2}{\xi^2}-1)}$ over the whole line space.

If let $t(\xi) = i\sqrt{\frac{1}{\xi^2}(\frac{\kappa^2}{\xi^2}-1)}$, $s(\xi) = 1$, $m(\xi) = \frac{1}{\xi^2}$, then $t(\xi) = h(m(\xi))$, where $h(\xi) = i\sqrt{\kappa^2\xi^2 - \xi}$ and we have

$$(\bar{T}_x)_{mn} = \int_{-1}^1\int_{-1}^1 \psi_m(x)\psi_n(y)\hat{t}(x-y)dxdy, \tag{6.1.38}$$

$$(\bar{S}_x)_{mn} = \int_{-1}^1\int_{-1}^1 \psi_m(x)\psi_n(y)\delta(x-y)dxdy, \tag{6.1.39}$$

$$(\bar{M}_x)_{mn} = \int_{-1}^1\int_{-1}^1 \psi_m(x)\psi_n(y)\hat{m}(x-y)dxdy. \tag{6.1.40}$$

Remark 6.1.3 : Although it is not clear if the system can apply the generalized eigenvalue diagonalization method or saying $\bar{S}_x, \bar{M}_x, \bar{T}_x$ are simultaneously diagonalizable, the numerical integration of the entries of $\bar{T}_x$ seems to be much easier due to the scaling effect of $\frac{1}{\xi^2}$. One interested study for the future could be to investigate the connection between $\bar{S}_x, \bar{M}_x, \bar{T}_x$.

6.1.4 Fourier Transform of Legendre Functions: Spherical Bessel Functions

The spherical Bessel functions j_n are defined by

$$j_n(x) = \sqrt{\frac{\pi}{2x}} J_{n+1/2}(x). \tag{6.1.41}$$

They can be connected with Legendre Polynomial from the following formula

$$\hat{L}_n(\xi) = \frac{1}{\sqrt{2\pi}} \int_{-1}^{1} L_n(x) e^{-i\xi x} dx = \frac{1}{\sqrt{2\pi}} \frac{1}{i^m} \sqrt{\frac{2\pi}{\xi}} J_{n+1/2}(\xi) = \frac{\sqrt{2}}{i^m \sqrt{\pi}} j_n(\xi), \quad \xi \in R, \tag{6.1.42}$$

Furthermore, $j_n(\xi)$ are real functions and

$$\int_R \hat{L}_n(\xi) \overline{\hat{L}_m(\xi)} d\xi = \frac{1}{\sqrt{2\pi}} \int_{-1}^{1} L_n(\xi) L_m(\xi) d\xi = \frac{1}{\sqrt{2\pi}} \frac{2}{2n+1} \delta_{mn}. \tag{6.1.43}$$

and

$$\begin{aligned}
&(\frac{1}{\xi^2} \hat{L}_n(\xi), \hat{L}_m(\xi)) \\
=& \int_{R_0} \frac{1}{\xi} \hat{L}_n(\xi)) \overline{\frac{1}{\xi} \hat{L}_m(\xi)} d\xi \\
=& \int_{R_0} (1/n \hat{L}'_n(\xi) + i/n \hat{L}_{n+1}(\xi)) \overline{(1/m \hat{L}'_m(\xi) + i/m \hat{L}_{m+1}(\xi))} d\xi \\
=& (x^2 L_n(x), L_m(x)) - (x L_n(x), L_{m+1}(x)) - (x L_{n+1}(x), L_m(x)) + (L_{n+1}(x), L_{m+1}(x))
\end{aligned} \tag{6.1.44}$$

which returns five-diagonal matrix by the orthogonality of the Legendre polynomials.

6.2 Spectral Methods for Non-linear Free Boundary Fluid Structure Interaction Problem.

This purpose of this work is to extend the numerical schemes developed in chapter 5 for the 2D non-linear no-slip boundary fluid structure interaction problem to the 3D free boundary fluid structure interaction problem. In this section, a brief idea on the numerical approximation for the 3D FSI problem is given. The numerical implantation and numerical analysis will be the future work.

The governing equations for this 3D FSI problem is:

$$\partial_t u - \Delta u + (u \cdot \nabla)u + \nabla p = 0, \qquad \Omega_f(t) \tag{6.2.1a}$$

$$\nabla \cdot u = 0, \qquad \Omega_f(t) \tag{6.2.1b}$$

$$\partial_{tt} w - \Delta w = 0, \qquad \Omega_s \tag{6.2.1c}$$

$$u = w_t, \qquad \Gamma_c(t)/\Gamma_c \tag{6.2.1d}$$

$$\partial_n w = \partial_n u - pn, \qquad \Gamma_c(t)/\Gamma_c \tag{6.2.1e}$$

where (u, p) posed in the Eulerian and w posed in the Lagrangian framework.

Assume position functions to be

$$\eta(\cdot, t) : \Omega_f \longrightarrow \Omega_f(t), \quad \eta(\cdot, t) : \Omega_s \longrightarrow \Omega_s(t)$$

then in the Lagrangian coordinate, we have

$$v(x,t) = \eta_t(x,t) = u(\eta(x,t),t), \qquad in \quad \Omega_f \tag{6.2.2a}$$

$$q(x,t) = p(\eta(x,t),t), \qquad in \quad \Omega_f \tag{6.2.2b}$$

$$w(x,t) = \eta(x,t) - x, \qquad in \quad \Omega_s. \tag{6.2.2c}$$

We obtain the new governing equations:

$$\partial_t v_i - \partial_j(a_{jl}a_{kl}\partial_k v_i) + \partial_k(a_{ki}q) = 0, \quad in \quad \Omega_f \qquad i = 1,2,3 \tag{6.2.3a}$$

$$a_{ki}\partial_k v_i = 0, \quad in \quad \Omega_f \tag{6.2.3b}$$

$$\partial_{tt} w_i - \Delta w_i = 0, \quad in \quad \Omega_s \qquad i = 1,2,3 \tag{6.2.3c}$$

with boundary condition on the common boundary Γ_c:

$$v_i = \partial_t w_i \qquad i = 1, 2, 3 \tag{6.2.4a}$$

$$\partial_j w_i N_j = a_{jl} a_{kl} \partial_k v_i N_j - q a_{ki} N_k \qquad i = 1, 2, 3 \tag{6.2.4b}$$

where the evolution of $a = a(x, t)$ defined for $x \in \Omega_f$, $t \geq 0$ by:

$$a(x, t) = (\nabla \eta(x, t))^{-1} \tag{6.2.5a}$$

$$a_{ij}(x, 0) = \delta_{ij} \tag{6.2.5b}$$

$$\partial_t a = -a : \nabla v : a \tag{6.2.5c}$$

where : denotes the matrix product.

For the well posedness of the above system, we refer to [29].

The rotational pressure correction scheme. For the time step $t_{n+1} = (n + 1)\triangle t$, the scheme is made up by the following three steps:

1. Compute a_{ij}^n based on v_i^n, η^n by an ODE solver on (6.2.5)

2. Given a_{ij}^n, v_i^n, q^n, compute $(\tilde{v}^{n+1}, w^{n+1})$ by

$$\frac{\tilde{v}_i^{n+1} - v^n}{\triangle t} - \partial_j (a_{jl}^n a_{kl}^n \partial_k \tilde{v}_i^{n+1}) + \partial_k (a_{ki}^n q^n) = 0, \; in \, \Omega_f \, i = 1, 2, 3 \tag{6.2.6a}$$

$$\partial_{tt} w_i^{n+1} - \Delta w_i^{n+1} = 0, \quad in \quad \Omega_s \quad i = 1, 2, 3 \tag{6.2.6b}$$

$$\tilde{v}_i^{n+1} = \frac{w_i^{n+1} - w^n}{\triangle t} \quad on \quad \Gamma_c \quad i = 1, 2, 3 \tag{6.2.6c}$$

$$\partial_j w_i^{n+1} N_j = a_{jl}^n a_{kl}^n \partial_k \tilde{v}_i^{n+1} N_j - q^n a_{ki}^n N_k \quad on \quad \Gamma_c \quad i = 1, 2, 3 \tag{6.2.6d}$$

,

3. Given $\tilde{v}^{n+1}$, update q^{n+1} and v^{n+1}.

$$\frac{v_i^{n+1} - \tilde{v}_i^{n+1}}{\triangle t} + \partial_k (a_{ki}^n (q^{n+1} - q^n + \lambda a_{mj}^n \partial_m \tilde{v}_j^{n+1})) = 0, \; i = 1, 2, 3 \tag{6.2.7a}$$

$$a_{kj}^n \partial_k v_j^{n+1} = 0, \quad in \quad \Omega_f \tag{6.2.7b}$$

$$\tilde{v}_i^{n+1} = 0, \quad q^{n+1} - q^n + \lambda a_{mj}^n \partial_m \tilde{v}_j^{n+1} = 0, \quad on \quad \Gamma_c \quad i = 1, 2, 3 \tag{6.2.7c}$$

.

Similarly, we can obtain an energy conservation property from the weak formulation, which has the form

$$\partial_t \left\{ \|u(t)\|^2_{\Omega_f} + \|w_t(t)\|^2_{\Omega_s} + \|\nabla w(t)\|^2_{\Omega_s} \right\} + 2\|a^T \nabla u(t)\|^2_{\Omega_f} = 0. \tag{6.2.8}$$

If define the system energy to be

$$E(u, w, w_t)(t) = \|u(t)\|^2_{\Omega_f} + \|w_t(t)\|^2_{\Omega_f} + \|\nabla w(t)\|^2_{\Omega_s}, \tag{6.2.9}$$

it can be showed that the system energy decays at the rate $2\|a^T \nabla u(t)\|^2_{\Omega_f}$.

Let $\delta u^n = \frac{u^n - u^{n-1}}{\triangle t}$ and define the discrete numerical energy E^n

$$E^n = \|u^n\| + \|\delta w^n\|^2 + \|\nabla w^n\|^2 + (\triangle t)^2 \|\nabla \cdot (a^n (p^n - \tilde{q}^n))\|^2. \tag{6.2.10}$$

Then we have the discrete numerical energy decay property:

$$\begin{aligned} & E^{n+1} - E^n \\ \leq & - \left\{ \|\tilde{u}^n - u^{n-1}\|^2 + (2 - d\lambda) \triangle t \|a^{nT} \nabla \tilde{u}^n\|^2 + \triangle t^2 \|\delta^2 w^n\|^2 + \triangle t^2 \|\delta \nabla w^n\|^2 \right\} \\ \leq & \quad 0. \end{aligned} \tag{6.2.11}$$

Therefore, the scheme proposed in this section is unconditional stable.

We leave the numerical implantation and further precise numerical analysis as our future work.

LIST OF REFERENCES

LIST OF REFERENCES

[1] L. Tsang, J. A. Kong, and R. T. Shin. *Theory of Microwave Remote Sensing*. Wiley, New York, 1985.

[2] P. G. Dinesen and J. S. Hesthaven. Fast and accurate modeling of waveguide grating couplers. *J. Opt. Soc. Am. A*, 17(9):1565–1572, 2000.

[3] P. G. Dinesen and J. S. Hesthaven. Fast and accurate modeling of waveguide grating couplers. ii. the three-dimensional vectorial case. *J. Opt. Soc. Am. A*, 18(11):2876–2885, 2001.

[4] L. C. Wilcox, P. G. Dinesen, and J. S. Hesthaven. Fast and accurate boundary variation method for multilayered diffraction optics. *J. Opt. Soc. Am. A*, 21(5):757–769, 2004.

[5] Mi-Sun Min, Tae-Woo Lee, Paul F. Fischer, and Stephen K. Gray. Fourier spectral simulations and gegenbauer reconstructions for electromagnetic waves in the presence of a metal nanoparticle. *Journal of Computational Physics*, 213(2):730–747, 2006.

[6] David Colton and Rainer Kress. *Inverse acoustic and electromagnetic scattering theory*. Springer-Verlag, Berlin, second edition, 1998.

[7] Oscar P. Bruno and Fernando Reitich. Numerical solution of diffraction problems: A method of variation of boundaries. *J. Opt. Soc. Am. A*, 10(6):1168–1175, 1993.

[8] Oscar P. Bruno and Fernando Reitich. Numerical solution of diffraction problems: A method of variation of boundaries. II. Finitely conducting gratings, Padé approximants, and singularities. *J. Opt. Soc. Am. A*, 10(11):2307–2316, 1993.

[9] Oscar P. Bruno and Fernando Reitich. Numerical solution of diffraction problems: A method of variation of boundaries. III. Doubly periodic gratings. *J. Opt. Soc. Am. A*, 10(12):2551–2562, 1993.

[10] D. Michael Milder. An improved formalism for rough-surface scattering of acoustic and electromagnetic waves. In *Proceedings of SPIE - The International Society for Optical Engineering (San Diego, 1991)*, volume 1558, pages 213–221. Int. Soc. for Optical Engineering, Bellingham, WA, 1991.

[11] D. Michael Milder. An improved formalism for wave scattering from rough surfaces. *J. Acoust. Soc. Am.*, 89(2):529–541, 1991.

[12] D. Michael Milder and H. Thomas Sharp. Efficient computation of rough surface scattering. In *Mathematical and numerical aspects of wave propagation phenomena (Strasbourg, 1991)*, pages 314–322. SIAM, Philadelphia, PA, 1991.

[13] D. Michael Milder and H. Thomas Sharp. An improved formalism for rough surface scattering. ii: Numerical trials in three dimensions. *J. Acoust. Soc. Am.*, 91(5):2620–2626, 1992.

[14] D. Michael Milder. Role of the admittance operator in rough-surface scattering. *J. Acoust. Soc. Am.*, 100(2):759–768, 1996.

[15] D. Michael Milder. An improved formalism for electromagnetic scattering from a perfectly conducting rough surface. *Radio Science*, 31(6):1369–1376, 1996.

[16] David P. Nicholls and Fernando Reitich. A new approach to analyticity of Dirichlet-Neumann operators. *Proc. Roy. Soc. Edinburgh Sect. A*, 131(6):1411–1433, 2001.

[17] David P. Nicholls and Fernando Reitich. Stability of high-order perturbative methods for the computation of Dirichlet-Neumann operators. *J. Comput. Phys.*, 170(1):276–298, 2001.

[18] David P. Nicholls and Fernando Reitich. Analytic continuation of Dirichlet-Neumann operators. *Numer. Math.*, 94(1):107–146, 2003.

[19] Ying He, David P. Nicholls, and Jie Shen. An efficient and stable spectral method for electromagnetic scattering from a layered periodic structure. *Journal of Computational Physics*, 231(8):3007–3022, 2012.

[20] J. A. Ogilvy. *Theory of wave scattering from random rough surfaces*. Adam Hilger Ltd., Bristol, 1991.

[21] Alexander G. Voronovich. *Wave scattering from rough surfaces*, volume 17 of *Springer Series on Wave Phenomena*. Springer-Verlag, Berlin, 1994.

[22] M. Saillard and A. Sentenac. Rigorous solutions for electromagnetic scattering from rough surfaces. *Waves Random Media*, 11(3):R103–R137, 2001.

[23] Karl F. Warnick and Weng Cho Chew. Numerical simulation methods for rough surface scattering. *Waves Random Media*, 11(1):R1–R30, 2001.

[24] J. A. DeSanto. *Scattering by Rough Surfaces*. In: Scattering, ed. by R. Pike, P. Sabatier. Academic Press, New York, 2002.

[25] Tanos Mikhael Elfouhaily and Charles-Antoine Guérin. A critical survey of approximate scattering wave theories from random rough surfaces. *Waves Random Media*, 14(4):R1–R40, 2004.

[26] S. K. Chakrabarti, S. Hernandez, and C. A. Brebbia, editors. *Fluid structure interaction and moving boundary problems*, volume 43 of *Advances in Fluid Mechanics*. WIT Press, Southampton, 2005. Edited papers from the 3rd International Conference on Fluid Structure Interaction and the 8th International Conference on Computational Modelling and Experimental Measurements of Free and Moving Boundary Problems held in La Coruna, September 19–21, 2005.

[27] Miguel A. Fernández. Coupling schemes for incompressible fluid-structure interaction: implicit, semi-implicit and explicit. *SẽMA J.*, (55):59–108, 2011.

[28] Gene Hou, Jin Wang, and Anita Layton. Numerical methods for fluid-structure interaction—a review. *Commun. Comput. Phys.*, 12(2):337–377, 2012.

[29] Mihaela Ignatova, Igor Kukavica, Irena Lasiecka, and Amjad Tuffaha. On well-posedness for a free boundary fluid-structure model. *J. Math. Phys.*, 53(11):115624, 13, 2012.

[30] Santiago Badia, Fabio Nobile, and Christian Vergara. Fluid-structure partitioned procedures based on Robin transmission conditions. *J. Comput. Phys.*, 227(14):7027–7051, 2008.

[31] Carlos A. Felippa, K.C. Park, and Charbel Farhat. Partitioned analysis of coupled mechanical systems. 190:32473270, 2001.

[32] Hermann G. Matthies and Jan Steindorf. Partitioned strong coupling algorithms for fluidstructure interaction. 81:805 – 812, 2003.

[33] Jaroslav Hron and Stefan Turek. *A monolithic FEM/multigrid solver for an ALE formulation of fluid-structure interaction with applications in biomechanics*. Springer, 2006.

[34] Björn Hübner, Elmar Walhorn, and Dieter Dinkler. A monolithic approach to fluid–structure interaction using space–time finite elements. *Computer methods in applied mechanics and engineering*, 193(23):2087–2104, 2004.

[35] Ulrich Küttler and Wolfgang A Wall. Fixed-point fluid–structure interaction solvers with dynamic relaxation. *Computational Mechanics*, 43(1):61–72, 2008.

[36] J. L. Guermond, P. Minev, and Jie Shen. An overview of projection methods for incompressible flows. *Comput. Methods Appl. Mech. Engrg.*, 195(44-47):6011–6045, 2006.

[37] Santiago Badia and Ramon Codina. On some fluid-structure iterative algorithms using pressure segregation methods. Application to aeroelasticity. *Internat. J. Numer. Methods Engrg.*, 72(1):46–71, 2007.

[38] L Tsang, J A Kong, and R T Shin. *Theory of Microwave Remote Sensing*. Wiley series in remote sensing. Wiley-Interscience, New York, 1985.

[39] L. Greengard and V. Rokhlin. A fast algorithm for particle simulations. *J. Comput. Phys.*, 73(2):325–348, 1987.

[40] D. Nicholls and Jie Shen. A stable, high–order method for two–dimensional bounded–obstacle scattering. *SIAM J. Sci. Comput.*, 28:1398–1419, 2006.

[41] Q. Fang, D. Nicholls, and Jie Shen. A stable, high–order method for two–dimensional bounded–obstacle scattering. *J. Comput. Phys.*, 224:1145–1169, 2007.

[42] David P. Nicholls and Jie Shen. A rigorous numerical analysis of the transformed field expansion method. *SIAM J. Numer. Anal.*, 47(4):2708–2734, 2009.

[43] F. Reitich and K. Tamma. State–of–the–art, trends, and directions in computational electromagnetics. *CMES Comput. Model. Eng. Sci.*, 5(4):287–294, 2004.

[44] George A. Baker, Jr. and Peter Graves-Morris. *Padé approximants*. Cambridge University Press, Cambridge, second edition, 1996.

[45] Carl M. Bender and Steven A. Orszag. *Advanced mathematical methods for scientists and engineers*. McGraw-Hill Book Co., New York, 1978. International Series in Pure and Applied Mathematics.

[46] Roger Petit, editor. *Electromagnetic theory of gratings*. Springer-Verlag, Berlin, 1980.

[47] M.O. Deville, P.F. Fischer, and E.H. Mund. *High-order methods for incompressible fluid flow*. Cambridge University Press, Cambridge, 500 p., 2002.

[48] Simon N. Chandler-Wilde and Bo Zhang. A uniqueness result for scattering by infinite rough surfaces. *SIAM J. Appl. Math.*, 58(6):1774–1790 (electronic), 1998.

[49] Simon N. Chandler-Wilde and Peter Monk. Existence, uniqueness, and variational methods for scattering by unbounded rough surfaces. *SIAM J. Math. Anal.*, 37(2):598–618, 2005.

[50] Simon N. Chandler-Wilde, Peter Monk, and Martin Thomas. The mathematics of scattering by unbounded, rough, inhomogeneous layers. *J. Comput. Appl. Math.*, 204(2):549–559, 2007.

[51] Armin Lechleiter and Sebastian Ritterbusch. A variational method for wave scattering from penetrable rough layers. *IMA J. Appl. Math.*, 75(3):366–391, 2010.

[52] Peijun Li and Jie Shen. Analysis of the scattering by an unbounded rough surface. *Math. Methods Appl. Sci.*, 35(18):2166–2184, 2012.

[53] Peijun Li, Haijun Wu, and Weiying Zheng. Electromagnetic scattering by unbounded rough surfaces. *SIAM J. Math. Anal.*, 43(3):1205–1231, 2011.

[54] Simon N. Chandler-Wilde, Eric Heinemeyer, and Roland Potthast. A well-posed integral equation formulation for three-dimensional rough surface scattering. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, 462(2076):3683–3705, 2006.

[55] Simon N. Chandler-Wilde, Chris R. Ross, and Bo Zhang. Scattering by infinite one-dimensional rough surfaces. *R. Soc. Lond. Proc. Ser. A Math. Phys. Eng. Sci.*, 455(1990):3767–3787, 1999.

[56] Bo Zhang and Simon N. Chandler-Wilde. Acoustic scattering by an inhomogeneous layer on a rigid plate. *SIAM J. Appl. Math.*, 58(6):1931–1950 (electronic), 1998.

[57] Bo Zhang and Simon N. Chandler-Wilde. Integral equation methods for scattering by infinite rough surfaces. *Math. Methods Appl. Sci.*, 26(6):463–488, 2003.

[58] J. A. DeSanto and P. A. Martin. On angular-spectrum representations for scattering by infinite rough surfaces. *Wave Motion*, 24(4):421–433, 1996.

[59] J. A. DeSanto and P. A. Martin. On the derivation of boundary integral equations for scattering by an infinite one-dimensional rough surfaces. *J. Acoust. Soc. Am.*, 102:67–77, 1997.

[60] J. A. DeSanto and P. A. Martin. On the derivation of boundary integral equations for scattering by an infinite two-dimensional rough surface. *J. Math. Phys.*, 39(2):894–912, 1998.

[61] S. Ritterbusch. Coercivity and the calderon operator on an unbounded domain. *PhD Thesis, Karlsruhe Institute of Technology*, 2009.

[62] Simon N. Chandler-Wilde and Johannes Elschner. Variational approach in weighted Sobolev spaces to scattering by unbounded rough surfaces. *SIAM J. Math. Anal.*, 42(6):2554–2580, 2010.

[63] Oscar P. Bruno and Fernando Reitich. Numerical solution of diffraction problems: A method of variation of boundaries; II. Finitely conducting gratings, Padé approximants, and singularities; III. Doubly periodic gratings. *J. Opt. Soc. Am. A*, 10(6, 11, 12):1168–1175, 2307–2316, 2551–2562, 1993.

[64] Oscar P. Bruno and Fernando Reitich. Boundary–variation solutions for bounded–obstacle scattering problems in three dimensions. *J. Acoust. Soc. Am.*, 104(5):2579–2583, 1998.

[65] Oscar P. Bruno and Fernando Reitich. High-order boundary perturbation methods. In *Mathematical Modeling in Optical Science*, volume 22, pages 71–109. SIAM, Philadelphia, PA, 2001. Frontiers in Applied Mathematics Series.

[66] David P. Nicholls and Fernando Reitich. Shape deformations in rough surface scattering: Cancellations, conditioning, and convergence. *J. Opt. Soc. Am. A*, 21:590–605, 2004.

[67] David P. Nicholls and Fernando Reitich. Shape deformations in rough surface scattering: Improved algorithms. *J. Opt. Soc. Am. A*, 21:606–621, 2004.

[68] David P. Nicholls and Jie Shen. A stable high-order method for two-dimensional bounded-obstacle scattering. *SIAM J. Sci. Comput.*, 28(4):1398–1419 (electronic), 2006.

[69] Qirong Fang, David P. Nicholls, and Jie Shen. A stable, high-order method for three-dimensional, bounded-obstacle, acoustic scattering. *J. Comput. Phys.*, 224(2):1145–1169, 2007.

[70] Ying He, David P. Nicholls, and Jie Shen. An efficient and stable spectral method for electromagnetic scattering from a layered periodic structure. *J. Comput. Phys.*, 231(8):3007–3022, 2012.

[71] J. Duoandikoetxea. Fourier analysis. *Graduate Studies in Mathematics*, 29, 2001.

[72] Jie Shen. Efficient spectral-Galerkin method. I. Direct solvers of second- and fourth-order equations using Legendre polynomials. *SIAM J. Sci. Comput.*, 15(6):1489–1505, 1994.

[73] J. Shen, T. Tang, and L. Wang. Spectral methods: algorithms, analysis and applications. *Springer Series in Computational Mathematics*, 41, 2011.

[74] Peijun Li. Coupling of finite element and boundary integral methods for electromagnetic scattering in a two-layered medium. *J. Comput. Phys.*, 229(2):481–497, 2010.

[75] C. Farhat, M. Lesoinne, and P. LeTallec. Load and motion transfer algorithms for fluid/structure interaction problems with non-matching discrete interfaces: momentum and energy conservation, optimal discretization and application to aeroelasticity. *Comput. Methods Appl. Mech. Engrg.*, 157(1-2):95–114, 1998.

[76] Erik Burman and Miguel A. Fernández. Stabilized explicit coupling for fluid-structure interaction using Nitsche's method. *C. R. Math. Acad. Sci. Paris*, 345(8):467–472, 2007.

[77] M. A. Fernández, J.-F. Gerbeau, and C. Grandmont. A projection semi-implicit scheme for the coupling of an elastic structure with an incompressible fluid. *Internat. J. Numer. Methods Engrg.*, 69(4):794–821, 2007.

[78] Santiago Badia, Annalisa Quaini, and Alfio Quarteroni. Splitting methods based on algebraic factorization for fluid-structure interaction. *SIAM J. Sci. Comput.*, 30(4):1778–1805, 2008.

[79] J. L. Guermond, P. Minev, and J. Shen. Error analysis of pressure-correction schemes for the time-dependent Stokes equations with open boundary conditions. *SIAM J. Numer. Anal.*, 43(1):239–258 (electronic), 2005.

[80] J. L. Guermond and Jie Shen. On the error estimates for the rotational pressure-correction projection methods. *Math. Comp.*, 73(248):1719–1737 (electronic), 2004.

[81] S. Dong and J. Shen. An unconditionally stable rotational velocity-correction scheme for incompressible flows. *J. Comput. Phys.*, 229(19):7013–7029, 2010.

[82] Jae-Hong Pyo. Error estimates for the second order semi-discrete stabilized gauge-Uzawa method for the Navier-Stokes equations. *Int. J. Numer. Anal. Model.*, 10(1):24–41, 2013.

[83] Q. Du, M. D. Gunzburger, L. S. Hou, and J. Lee. Semidiscrete finite element approximations of a linear fluid-structure interaction problem. *SIAM J. Numer. Anal.*, 42(1):1–29, 2004.

[84] Igor Kukavica, Amjad Tuffaha, and Mohammed Ziane. Strong solutions to a nonlinear fluid structure interaction system. *J. Differential Equations*, 247(5):1452–1478, 2009.

[85] J. L. Guermond and Jie Shen. Velocity-correction projection methods for incompressible flows. *SIAM J. Numer. Anal.*, 41(1):112–134 (electronic), 2003.

[86] Alfio Quarteroni and Alberto Valli. *Domain decomposition methods for partial differential equations*. Numerical Mathematics and Scientific Computation. The Clarendon Press Oxford University Press, New York, 1999. Oxford Science Publications.

[87] Andrea Toselli and Olof Widlund. *Domain decomposition methods—algorithms and theory*, volume 34 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, 2005.

[88] Ying He, David P. Nicholls, and Jie Shen. An efficient and stable spectral method for electromagnetic scattering from a layered periodic structure. *J. Comput. Phys.*, 231(8):3007–3022, 2012.

[89] Jie Shen. Efficient spectral-Galerkin method. I. Direct solvers of second- and fourth-order equations using Legendre polynomials. *SIAM J. Sci. Comput.*, 15(6):1489–1505, 1994.

[90] Feng Chen and Jie Shen. A GPU parallelized spectral method for elliptic equations in rectangular domains. *Journal of Computational Physics*, 250:555–564, October 2013.

VITA

VITA

Ying He is a Mathematics Ph.D. student in the Department of Mathematics at Purdue University, West Lafayette. Her thesis advisor is Professor Jie Shen. She worked on her Ph.D. degree in applied mathematics since 2008. Before that, she received her M.S. degree (2008) and B.S. degree (2005) in Applied Mathematics both from the School of Mathematical Sciences at Xiamen University in Xiamen, P.R. China. In May 2013, she also received her M.S. degree in Computer Science at Purdue University, West Lafayette.

Ying He's research interest lies in applied and computational mathematics, in particular numerical analysis and scientific computing with applications in acoustic and electromagnetic scattering, fluid dynamics, and parallel algorithms for scientific computing.

During January 2010 – May 2013, Ying He worked as a research assistant student at Purdue University, under the supervision of Prof. Jie Shen. She focused on developing high order numerical methods for acoustic and electromagnetic scattering problems, and nonlinear fluid-structure interaction problems.

During May – August, 2012 and May – August, 2013, Ying He was a Givens Associate summer internship student at Argonne National Laboratory (ANL), under the supervision of Dr. Misun Min, computational scientist, MCS Division, ANL.

Ying He's publications are attached below.

Publication List:

1. Y. He, D. P. Nicholls, and J. Shen. *An efficient and stable spectral method for electromagnetic scattering from a layered periodic structure.* J. Comput. Phys., 231(8):3007–3022, 2012.
2. Y. He, P. Li, and J. Shen. *A New Spectral Method for Numerical Solution of the Unbounded Rough Surface Scattering Problem.* submitted, 2013.
3. (Preprint, with Misun Min and David P. Nicholls) *A Spectral Element Method with Transparent Boundary Conditions for Acoustic Time-Harmonic Scattering in Quasi-Periodic Double-Layer Structures.*
4. (Preprint, with Jie Shen) *Unconditional Stable Pressure-Correction Schemes for Non-linear No-slip Fluid Structure Interaction Problem.*
5. (In preparation, with Jie Shen) *Spectral Methods for Non-linear Free Boundary Fluid Structure Interaction Problem.*